

Szegedi Tudományegyetem  
Természettudományi és Informatikai Kar  
Biológia Doktori Iskola

# **CRISPR/Cas rendszerek és kis RNS-eik expressziójára alkalmazott promóterek vizsgálata**

Doktori (Ph.D.) értekezés

**Weinhardt Nóra**

Témavezető: Welker Ervin PhD, DSc

Biokémiai Intézet  
Szegedi Biológiai Kutatóközpont  
Magyar Tudományos Akadémia

Enzimológiai Intézet  
Természettudományi Kutatóközpont  
Magyar Tudományos Akadémia



2020  
Szeged

## TARTALOMJEGYZÉK

1. Rövidítések jegyzéke.....	4
2. Bevezetés.....	5
2.1. CRISPR áttekintés .....	5
2.2. A CRISPR/Cas rendszerek off-target hatása és az ennek csökkentésére irányuló törekvések.....	11
2.3. Cas12a nukleázok .....	13
2.4. A CRISPR/Cas9 működését befolyásolja a gRNS-ek mennyisége és hossza.....	14
2.5. A CRISPR gRNS-ek expressziójára alkalmazott promóterek.....	15
3. Célkitűzések .....	20
4. Anyagok és módszerek.....	22
4.1. Plazmidok létrehozása a Cas12a nukleázok és a megnövelt specificitású SpCas9 variációk vizsgálatához.....	22
4.2. Sejttenyésztés.....	23
4.3. A GFxFP vizsgálat.....	23
4.4. Genomi homológ rekombinációs teszt .....	24
4.5. EGFP hasítási teszt .....	24
4.6. Áramlási citometria .....	25
4.7. Plazmid könyvtárak létrehozása .....	26
4.7.1. Plazmid könyvtár vektorainak létrehozása .....	27
4.7.2. Plazmid könyvtár inszertek létrehozása.....	28
4.7.3. A vektorok és inszertek összeillesztése .....	28
4.7.4. Transzformáció .....	28
4.7.5. Plazmid könyvtárak tisztítása.....	29
4.8. A cDNS könyvtárak létrehozása.....	29
4.8.1. Emlős sejtek RNS könyvtárainak létrehozása .....	29
4.8.2. Bakteriális promóteres RNS könyvtárak létrehozása.....	29
4.8.3. In vitro T7 transzkripció .....	30
4.8.4. Reverz transzkripció .....	30
4.9. Adapterek ligálása a cDNS-re .....	30
4.10. NGS .....	31
4.11. Programozás .....	31
4.12. Statisztika.....	31
5. Eredmények.....	33
5.1. Cas12a nukleázok vizsgálata .....	33
5.1.1. A GFxFP esszé módosítása.....	33
5.1.2. Az As- és LbCas12a hatékonyan indukálja a GFP kazetta javítását a GFxFP esszében .....	36
5.1.3. Az As- és LbCas12a nukleázok által végzett genomi DNS-hasítás hatékony HR javítást képes indukálni.....	38

5.2. Megnövelt specificitású SpCas9 nukleázok vizsgálata .....	40
5.2.1. 5' végükön nem G-t, vagy a 20. pozícióban nem komplementer G-t tartalmazó gRNS-ek alkalmazása .....	41
5.2.2. 5' végükön rövidebb gRNS-ek alkalmazása .....	42
5.3. A CRISPR gRNS-ek expressziójára gyakran alkalmazott promóterek .....	43
5.3.1. A randomizált bázisokkal módosított gRNS .....	43
5.3.2. cDNS+adapter ligálási módszer .....	44
5.3.3. Emlős promóterek összehasonlítása .....	45
5.3.4. A promótortól downstream elhelyezkedő +1-től +4-ig terjedő pozíciók vizsgálata az RNS expresszióra és a transzkripció iniciációs pozícióra .....	46
5.3.4.1. Az U6 promóter .....	48
5.3.4.2. A 7sk promóter .....	55
5.3.4.3. A T7 promóter .....	63
5.3.4.4. Az U6T7 promóter .....	72
5.3.4.5. A J23119 promóter .....	77
6. Diszkusszió .....	85
6.1. A Cas12a nukleázok .....	85
6.2. A megnövelt specificitású SpCas9 variánsok .....	86
6.3. A CRISPR gRNS-ek expressziójára gyakran alkalmazott promóterek .....	87
6.3.1. Az U6 promóter .....	87
6.3.2. A 7sk promóter .....	88
6.3.3. A T7 promóter .....	89
6.3.4. Az U6T7 promóter .....	89
6.3.5. A J23119 promóter .....	90
6.3.6. Konklúzió .....	91
7. Köszönetnyilvánítás .....	93
8. Közreműködések .....	94
9. Irodalomjegyzék .....	95
10. Összefoglalás .....	98
11. Summary .....	103
12. Publikációs lista .....	108
13. Függelék .....	109

## 1. Rövidítések jegyzéke

+1A	+1-es pozícióban adenin nukleotid
+1C	+1-es pozícióban citozin nukleotid
+1G	+1-es pozícióban guanin nukleotid
+1T	+1-es pozícióban timin nukleotid
A	Adenin
AsCas12a	<i>Acidaminococcus</i> sp. BV3L6 Cas12a
ATP	Adenozin-trifoszfát
C	Citozin
Cas	CRISPR asszociált fehérje (CRISPR associated protein)
cDNS	Komplementer DNS
Chl	Kloramfenikol
Cpf1	<i>Prevotella</i> és <i>Francisella</i> 1 CRISPR
CRISPR	Halmazottan szabályos közökkel elválasztott rövid palindromikus ismétlődések (Clustered Regularly Interspaced Short Palindromic Repeats)
CRISPRa	CRISPR aktiválás
CRISPRi	CRISPR interferencia
crRNA	CRISPR RNS
dCas	Nukleáz inaktivált Cas fehérje (dead Cas protein)
dead Cas	Nukleáz inaktivált Cas fehérje
DMEM	Dulbecco's Modified Eagle's Médium
DNS	Dezoxiribonukleinsav
DSE	Disztális szekvencia elem
<i>E. coli</i>	<i>Escherichia Coli</i>
EGFP	Megnövelt Zöld Fluoreszcens fehérje (Enhanced Green Fluorescent Protein)
indel	Inszerció, deléció mutációk
iRFP	Infravörös fluoreszcens fehérje
G	Guanin
GFP	Zöld fluoreszcens fehérje
gRNS	Vezető RNS (guide RNA)
guideRNS	Vezető RNS (guide RNA)
HR	Homológ rekombinációs javítás
LB	Lisogeny broth
LbCas12a	<i>Lachnospiraceae</i> ND2006 Cas12a
MCS	Multiklónozó hely
N2a	Neuro 2-a
NGS	Új-generációs szekvenálás (Next-Generation Sequencing)
NHEJ	Nem homológ végek összeillesztése (Non-Homologous End Joining)
NmCas9	<i>Neisseria meningitidis</i> Cas9
PAM	Protospacer melletti szekvencia (Protospacer-Adjacent Motif)
PCR	Polimeráz láncreakció
PEG	Polietilén-glikol
PolIII	Polimeráz III
PSE	proximális szekvencia elem
RNAP	RNS polimeráz
RNP	Ribonukleoprotein (Ribonucleoprotein)
RNS	Ribonukleinsav
RT	Reverz transzkriptáz
SaCas9	<i>Staphylococcus aureus</i> Cas9
SpCas9	<i>Streptococcus pyogenes</i> Cas 9
SSA	Egyszálú kapcsolódás (Single strand annealing)
StCas9	<i>Streptococcus thermophilus</i> Cas9
T	Timin
TALEN	Transzkripció aktivátor-szerű effektor nukleáz (Transcription Activator-Like Effector Nuclease)
tracrNS	Transz-aktiváló crRNS (trans-activating crRNA)
ORI	Replikációs origó
ZFN	Cink-Ujj nukleáz (Zinc Finger Nuclease)

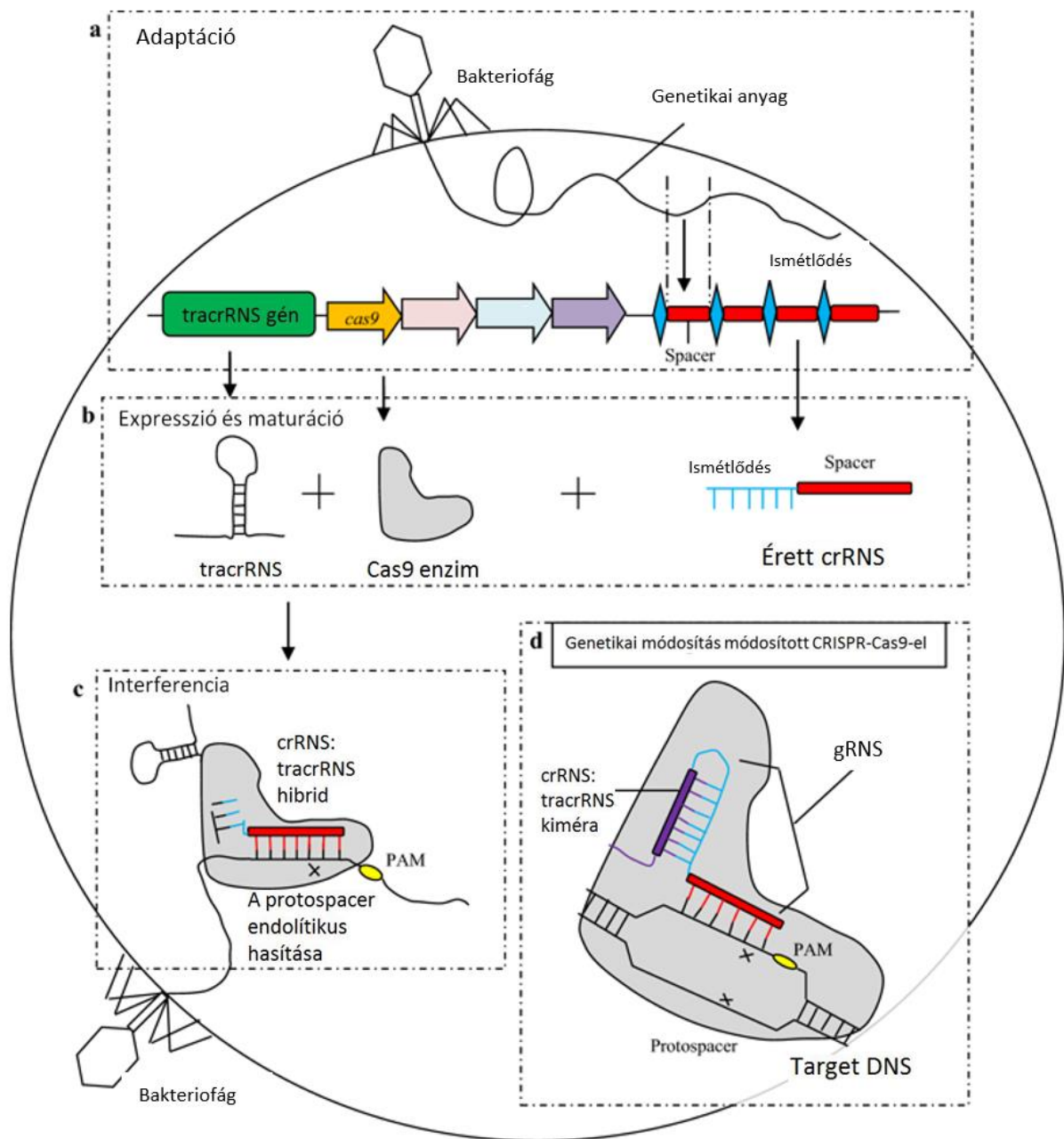
## 2. Bevezetés

### 2.1. CRISPR áttekintés

Számtalan biológiai mechanizmus felfedezéséhez és tanulmányozásához elengedhetetlen a genetikai információ szerkesztésének képessége. Az alkalmazott tudományok, a biotechnológia és az orvostudomány területén is hatalmas szükség van olyan technikákra, amelyekkel a DNS specifikus szakaszait lehet irányíthatóan módosítani. A mindössze néhány éve elterjedt CRISPR rendszer forradalmi változásokat okozott annak köszönhetően, hogy egyszerűen, egy rövid RNS szekvencia módosításával képes nagy hatékonysággal a kívánt DNS szakaszt módosítani.

A CRISPR rendszer elterjedését megelőzően restrikciós enzimeken alapuló génszerkesztő technológiákat is alkalmaztak a genom módosítására, ilyenek a cink-ujj nukleázok (ZFN) és az aktivátorszerű effektor nukleázok (TALEN). Ezen technológiák hátránya, hogy minden egyes kísérlethez fehérjeszinten kell megváltoztatni a molekulákat, mivel ezeknél a módszereknél a DNS szekvenciák felismerése nem RNS-DNS, hanem fehérje-DNS kötésen alapul<sup>1,2</sup>.

A természetben a CRISPR/Cas rendszer a külső behatolókkal szembeni védekezés része prokariótákban, melynek során felismerik és elhasítják a támadó organizmusok nukleinsavait. A védekezés az adaptáció, az expresszió és érés, valamint az interferencia szakaszokból áll (1. ábra). Az adaptáció során a Cas9 fehérjék az idegen DNS-ből rövid szakaszokat (a protospacereknek nevezett szekvenciákat) a prokarióta genomba építik. Ezeket a beépített szekvenciákat spacereknek hívják (1. ábra, a). A spacereket a CRISPR lókuszt ismétlődő szekvenciái határolják, és egy újabb támadás esetén felhasználhatók a védekezéshez. Az expressziós és érési szakaszban átíródnak a Cas9 fehérjét, a pre-crRNS-eket (CRISPR RNS), és a tracrRNS-t (transzaktiváló crRNS) kódoló gének. A pre-crRNS-ekből létrejönnek az egy-egy spacer szekvenciát tartalmazó, egyedi crRNS-ek (1. ábra, b). Az interferencia fázisban létrejön a Cas9-crRNS és tracrRNS fehérje-RNS komplex. A tracrRNS részben komplementer a crRNS-el és együttesen egy RNS duplexet hoznak létre, amelyet felismer és köt a Cas9. A nukleáz a spacer szekvencia segítségével Watson-Crick bázispárosodás alapján azonosítja az idegen DNS molekulákat, amelyeket mindkét szálon elhasít (1. ábra, c). A cél-DNS felismeréséhez szükség van még egy konzervált, néhány bázispáros szakasz jelenlétére a célszekvencia mellett, ez a protospacer melletti motívum (PAM) szekvencia, amelynek segítségével a Cas9-crRNS-tracrRNS komplex megkülönbözteti az idegen DNS-t a saját DNS-től<sup>2-4</sup>.



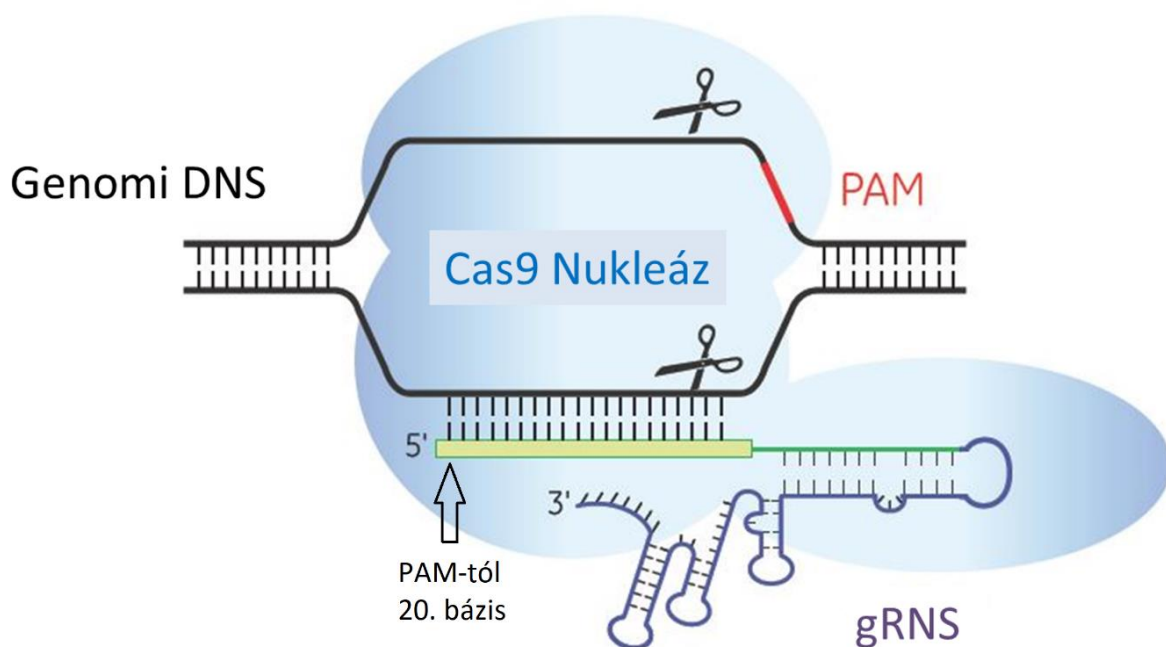
**1. Ábra. A prokariótákban létező természetes CRISPR/Cas9 rendszer mechanizmusa és a genomszerkesztésre használt módosított CRISPR/Cas9 technológia**

**a – c,** A természetben létező CRISPR/Cas9 rendszer mechanizmusának három fő fázisa van.

**a,** Adaptáció: a baktériumok az őket támadó vírusokból a „protospacernek” nevezett specifikus genomi szekvenciákat szereznek, és beépítik azokat a CRISPR lókuszbba. Ezek a spacerok. **b,** Expresszió és érés: A tracrRNA, a Cas9 enzim és a crRNA expresszálódnak. A pre-crRNA-ekből egyedi spacerokat tartalmazó crRNA-ek keletkeznek. **c,** Interferencia: megtörténik a crRNA és a tracrRNA bázispárosodása, amely crRNA:tracrRNA hibriddé alakul, majd ezek később a Cas9 fehérjével komplexet alkotnak. A hibrid RNA alapján a Cas9 fehérje megtalálja a spacerrel komplementer protospacer célszekvenciát, melyek között bázispárosodás történik, majd a Cas9 nukleáz hasítja a target DNS mindkét szálát. **d,** A genomszerkesztés során a kiméra gRNA-t a tracrRNA és a crRNA hibridizációjával állítják elő. A gRNA spacere alapján a Cas9 nukleáz felismeri azokat a célszekvenciákat, amelyek a protospacerrel szomszédos motívum (PAM) szekvenciától upstream helyezkednek el és a DNS mindkét szálát hasítja<sup>2</sup>. Ábra forrása: Barman és mtsai, 2019 alapján.

A *Streptococcus pyogenes* baktériumból származó Cas9 endonukleáz (SpCas9) jelenleg a génmódosításra legelterjedtebben alkalmazott CRISPR/Cas rendszer. A specificitást a crRNS (1. ábra, c), vagy ennek mesterségesen továbbfejlesztett verziójaként a crRNS és a tracrRNS fúziójából létrehozott guideRNS (gRNS) 5' végén elhelyezkedő 20 bázispáros szekvencia, a spacer biztosítja. A spacer segítségével kerül felismerésre a célszekvencia (target) komplementer DNS-RNS hibridizáció alapján (1. ábra d, és 2. ábra). A target szekvencia duplaszálú DNS-e kihurkolódik és a komplementer szekvenciák között RNS-DNS heteroduplex keletkezik a gRNS és a cél DNS között. A Cas9 nukleáz hasítja a targetet és a targettel komplementer DNS szálát is, így kettős szálú töréseket generál<sup>2</sup>.

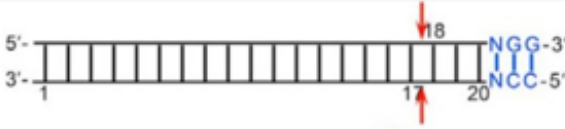
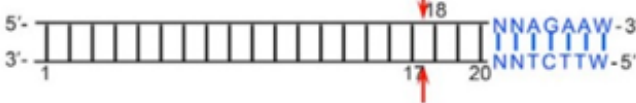


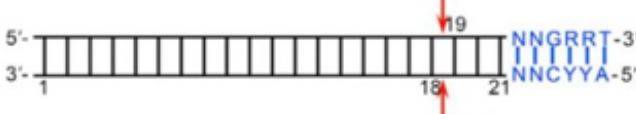
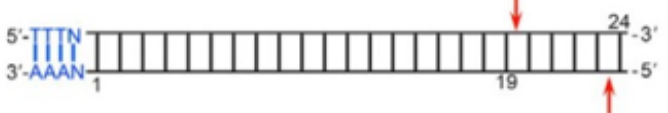
A mai molekuláris biológiai technikákkal már könnyedén változtatható a spacer szekvenciája, ennek köszönhető a rendszer robbanásszerű elterjedése<sup>2,5</sup>. A Cas9 fehérje segítségével szinte bármilyen genomi szakasz szabadon módosítható, ezért jelentős biotechnológiai, orvosi és tudományos felhasználása van.



## 2. Ábra. Cas9 nukleáz a gRNS-el és a genomi target DNS-el

Az ábrán a Cas9 nukleáz, a gRNS és a genomi target DNS együttesen látható sematikus ábrázolva. Kék színnel a fehérje, ollókkal a DNS duplaszálú hasítása, feketével a target és környezetének szekvenciája, pirossal a PAM szekvencia van jelölve. A gRNS spacer szekvenciája, amely a targettel komplementer, sárgával van jelölve és a guideRNS 3' részével folytatódik, amelynek erős másodlagos szerkezete van. A spacerben fekete nyíllal van jelölve a gRNS 5' első bázisa, amelyet 20. pozíciónak szoktak nevezni, mert a PAM-tól 20 bázisra található a targetben az ennek megfelelő nukleotid. Ábra forrása: <https://horizondiscovery.com/en/products/gene-editing/gene-editing-reagents/crispr-guide-rna> oldal alapján.

Az SpCas9 az NGG szekvenciájú, 3 bázispáros PAM motívumot ismeri fel. A DNS duplaszáltörése az SpCas9 esetében a targetben a 17. és a 18. bázispárok között történik a spacer 5' végétől számolva és tompa végek keletkeznek (3. ábra). A Bacteriák és Archeák különböző fajai között a CRISPR/Cas rendszer számos variációja alakult ki, amelyek alapján a CRISPR rendszerek többféle osztályba és csoportba sorolhatók. A különböző, eddig felfedezett CRISPR/Cas fehérjék eltérő PAM motívumokat ismerhetnek fel, és eltérő lehet a DNS hasítási módja is (3. ábra)<sup>6</sup>.

Enzim neve	Mérete	PAM motívumja és DNS hasítási módja
SpCas9 / FnCas9	1368 / 1629	
St1Cas9	1121	
St3Cas9	1409	
NmCas9	1082	
SaCas9	1053	
AsCpf1 / LbCpf1	1307 / 1228	

### 3. Ábra. Néhány eddig felfedezett CRISPR/Cas rendszer PAM szekvenciája és DNS hasítási módja

Az ábrán néhány eddig felfedezett Cas fehérje és az általuk felismert PAM motívumok, valamint a DNS hasítási módjuk látható. Kék színnel a PAM motívumok szekvenciái, piros nyilakkal a PAM szekvenciától a megjelölt bázispárnai távolságra létrejövő hasítás helye látható. <sup>6</sup>. Ábra forrása: Komor és mtsai, 2016 alapján.

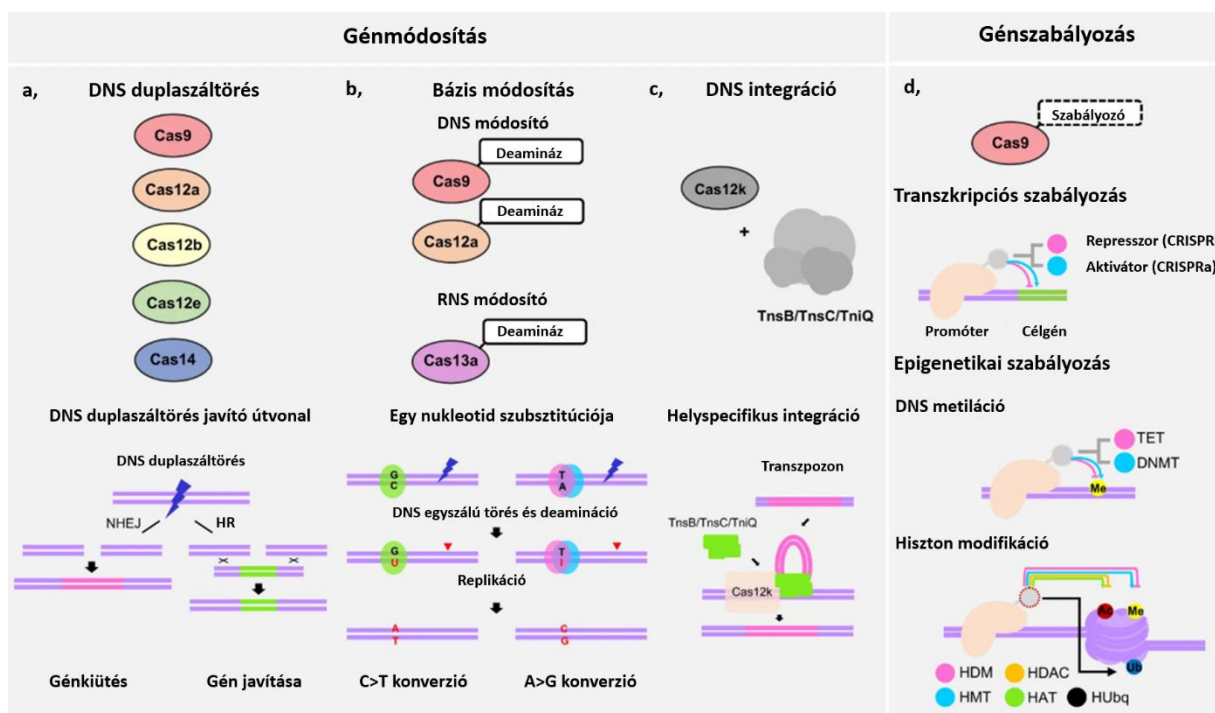
A különféle Cas nukleázok által létrehozott DNS duplaszáltörés két fő mechanizmussal javítható a sejtekben, és mindkét javítási mechanizmus által módosulhat a DNS szekvencia (4. ábra, a). Az NHEJ mechanizmus által indelt mutációk hozhatók létre a kívánt szakaszon, a homológ rekombinációs hibajavítás (HR) során pedig templát DNS jelenlétében a szekvencia pontos módosítása érhető el<sup>7</sup>. Az SSA hibajavítás a homológ rekombinációs javítás egyik altípusa (nincs ábrázolva), amely két ismétlődő szekvencia között történik.



Az SSA hibajavítás során az 5-3' irányban visszaemésztett szálaknál a két ismétlődő szekvencia szolgál templátul, amelynek eredményeképp az eredetileg kétszer előforduló szekvencia helyett csak 1 kópia marad meg és a közbenső szekvenciák kiesnek.

A CRISPR rendszerrel többféle módon valósítható meg a genom irányított szerkesztése, melynek néhány típusa a 4. ábra a, b és c részén látható. Génmódosítás hozható létre „klasszikus” módon, a duplaszálú DNS hasítása által kiváltott DNS javítási eljárással (3. ábra a). A DNS kötésre képes, de hasításra nem képes inaktív dCas (nukleáz inaktivált, dead Cas) fuzionálható deaminázokkal, hogy egy aminocsoportot a citozinban, vagy adeninben hidrolizáljanak, mellyel specifikus bázis módosítás, C->T, illetve A->G konverzió érhető el. Az RNS-ek bázis szerkesztéséhez az RNS-re ható adenzin-dezaminázokat fuzionálják a dCas13 enzimhez<sup>7-9</sup> (4. ábra b). Egy másik módszer a transzpozázilis elemek helyspecifikus integrációját képes indukálni a transzpozáz komplexek inaktív Cas12-vel történő irányításával<sup>10</sup> (4. ábra c). A prime editing nevű módszer szintén alkalmas génszerkesztésre (nincs ábrázolva). Ennek során egy, csak a DNS egyik szálát hasító mutáns Cas9-et (nikáz) fuzionálnak reverz transzkriptázzal. A gRNS 3' végét pedig a target módosítására alkalmas donor szekvenciával hosszabbítják meg, amelyet a reverz transzkriptáz átír, a keletkezett ssDNS pedig beépül az egyik szálon hasított target szekvenciába<sup>11</sup>.

A CRISPR/Cas rendszerrel génszabályozás is végbe vihető. Többféle transzkripció, vagy epigenetikus szabályozót lehet közvetlenül fuzionálni, vagy RNS-protein kölcsönhatásokkal közvetve kapcsolni a dCas-okhoz. A gének expresszióját lehet aktiválni, (CRISPRa), vagy gátolni (CRISPRi), a kapcsolt szabályozó fehérjék funkcionális jellegétől függően. Helyspecifikus génszabályozás hozható létre még epigenetikai szabályozók fúziójával szintén a dCas nukleázok segítségével (4. ábra d).



#### 4. Ábra. Génmódosítás és génszabályozás CRISPR-el

Az ábrán a CRISPR/Cas rendszerek által elérhető főbb génmódosítási és génszabályozási típusok láthatók. **a**, A különféle nukleázok által létrehozott DNS duplaszáltörés két fő mechanizmussal javítható a sejtekben, NHEJ, vagy homológ rekombinációs hibajavítással (HR). **b**, Az inaktív Cas nukleázok fuzionálhatók deaminázokkal, mellyel specifikus bázis módosítás érhető el. Az RNS-ek bázis szerkesztéséhez az RNS-re ható adenzin-dezaminázokat fuzionálják a dCas13 enzimhez. **c**, A transzpozázilis elemek helyspecifikus integrációját képesek indukálni a transzpozáz komplexek, inaktív Cas12 irányításával. **d**, A transzkripció, vagy epigenetikus szabályozókat lehet közvetlenül fuzionálni, vagy RNS-protein kölcsönhatásokkal közvetve kapcsolni a dCas-okhoz. A gének expresszióját lehet aktiválni, (CRISPRa), vagy gátolni (CRISPRi), a kapcsolt szabályozó fehérjék funkcionális jellegétől függően. Génszabályozás érhető el epigenetikai szabályozók fúziójával a dCas nukleázokon keresztül<sup>7</sup>. Ábra forrása: Moon és mtsai 2019 alapján.

Annak ellenére, hogy a CRISPR/Cas rendszereknek ma már számtalan biotechnológiai alkalmazása van, vannak limitációi is. Az egyik, hogy nem minden target szekvenciával képesek hatékony működésre. Ennek megoldására nyújthat segítséget a többféle ismert nukleáz alkalmazása, valamint olyan Cas9 verziók használata, ahol az SpCas9 nemcsak NGG, hanem NG PAM-al is képes működni, amelyekkel megnövelhető a lehetséges targetek száma<sup>12,13</sup>. A CRISPR-es alkalmazásoknál nagy nehézséget okoz a hatékony bejuttatás is, ezeknek a fejlesztésével számos csoport foglalkozik<sup>14</sup>. A CRISPR másik hátránya az off-target hatás, amely részletesebben a következő fejezetben kerül kifejtésre.

## 2.2. A CRISPR/Cas rendszerek off-target hatása és az ennek csökkentésére irányuló törekvések

A CRISPR/Cas rendszerek használata során felmerülő egyik legnagyobb probléma az off-target hatás. A gyakorlatban ez azt jelenti, hogy a CRISPR/Cas rendszerek által célozni kívánt DNS szekvencián kívül a Cas fehérjék más, a célszekvenciához hasonló genomi pozíciókat is hasítanak<sup>15</sup>. Ez nemkívánt mellékhatásokat eredményezhet a kísérlet, vagy a potenciális gyógyászati beavatkozás során.

Ma már a különböző Cas fehérjékre többféle adatbázis nyújthat segítséget olyan targetek kiválasztásához, amelyeknek viszonylag kevés számú off-targete van az adott organizmusban. Azonban az változatlanul rosszul prediktálható, hogy melyik off-target kerül hasításra. Ezért számos stratégiát közöltek a szakirodalomban a gRNS módosításától kezdve, a Cas9 fehérjék módosításán keresztül a fúziós fehérje rendszerek létrehozásáig<sup>15</sup>.

Az 5. ábra a részén szerepelnek ezek közül a megnövelt specificitású Cas9 variációk időrendi sorrendben az SpCas9<sup>2Pro</sup>-ig bezárólag. Az egyik legizgalmasabb és egyben legelterjedtebb módszer ezen mutációkkal létrehozott megnövelt specificitású SpCas9 variánsok alkalmazása az off-target hatás csökkentése érdekében. Az eSpCas9<sup>16</sup>, SpCas9-HF1<sup>17</sup>, HypaCas9<sup>18</sup> és SpCas9<sup>2Pro</sup><sup>19</sup> a fehérjeszerkezet ismeretében, különböző mutációk bevitelével készültek, az evoCas9<sup>20</sup>, xCas9<sup>12</sup> és a Sniper-Cas9<sup>21</sup> pedig irányított evolúcióval (5. ábra a).

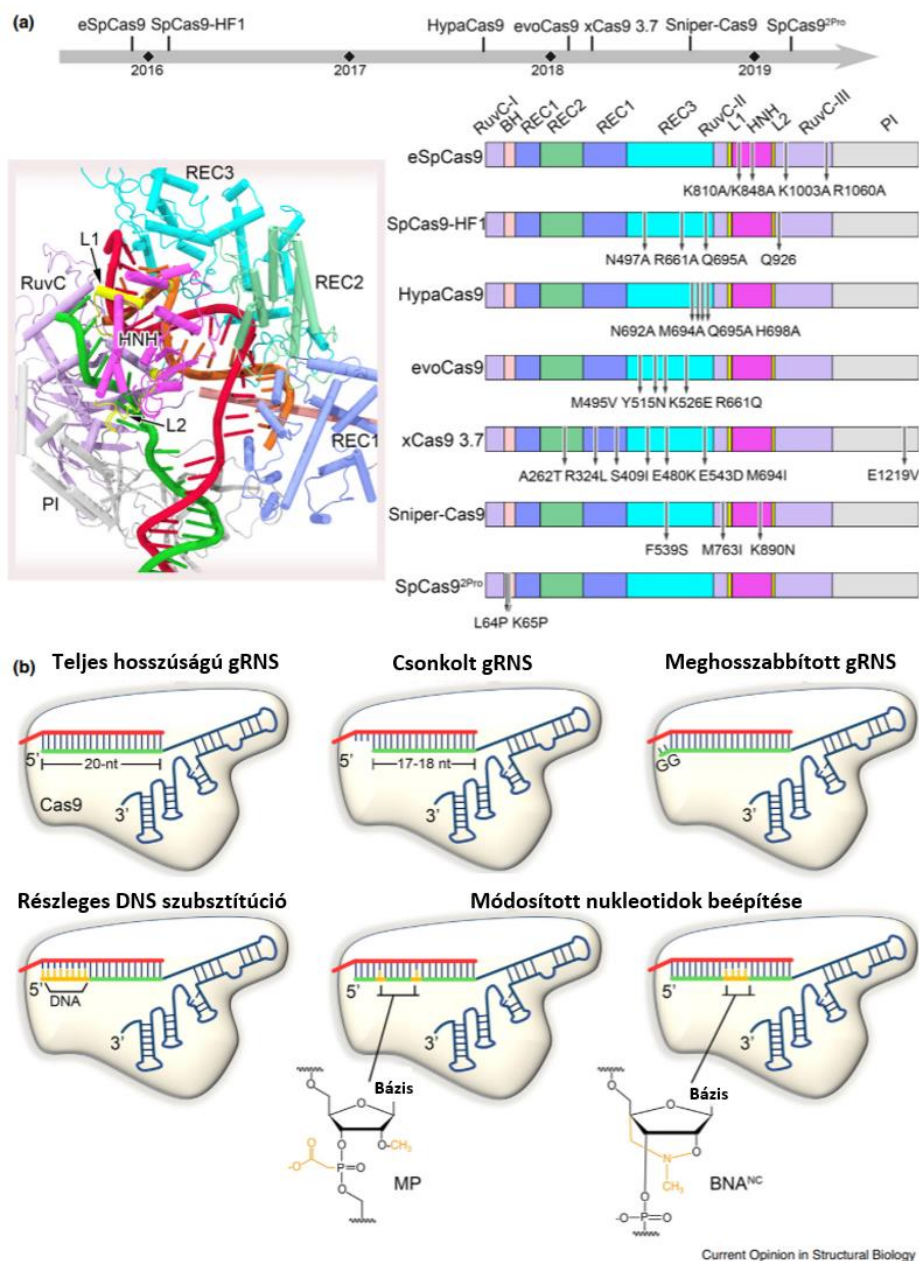
Az off-target hatás csökkentésére irányuló másik módszer a gRNS-ek módosítása. Ezek többsége az RNS 5' végén fordul elő, beleértve a gRNS kettő vagy három bázissal való csonkítását<sup>22</sup>, két guanin nukleotiddal való meghosszabbítását<sup>23</sup> és az 5' vég DNS-el való helyettesítését<sup>15,24</sup>. A kémiaiailag módosított nukleotidok, például a 2'-O-metil-3'-foszfonoacetát (MP) és az áthidaló nukleinsavak (2', 4'-BNANC [N-Me]) a gRNS meghatározott szakaszain szintén nagymértékben csökkentik az off-target hatást<sup>15,25,26</sup> (5. ábra b).

Az 5. ábrán felsorolt megoldásokon kívül elterjedt módszernek számít még a dCas9 fúziója különböző fehérjedoménokkal. Ilyen például az inaktivált dCas9-el fuzionált FokI nukleázok alkalmazása<sup>27</sup>. Mivel a dCas9 csak a DNS kötésére képes, semmilyen targetet nem képes hasítani, így az off-targeteket sem. A fúziós megoldások közül az egyik legújabb megközelítés a dCas9 és egy bázis módosító fehérje (mint a citidin vagy adenin deamináz) fúziója, amelynek segítségével DNS törése nélkül is lehet mutációkat bevinni a génekbe (4. ábra, b)<sup>28-30</sup>.

Egy másik módszer az off-target hatás csökkentésére a Cas nukleáz aktivitásának időbeli csökkentése, ami például indukálható nukleázokkal érhető el. A nukleáz aktivitási idejének csökkentésével a kívánt targetek még elhasítódnak, az off-targetek pedig már kevésbé<sup>31,32</sup>.

Az off-target hatást sikeresen csökkenti még a nikázok használata. Ezek olyan mutáns nukleázok, amelyek csak a DNS egyik szálát képesek hasítani<sup>33</sup>.

A felsorolt módszereken kívül többen bemutatták, hogy a Cas12a nukleázok specifikusabbak, mint az SpCas9<sup>34-36</sup>.

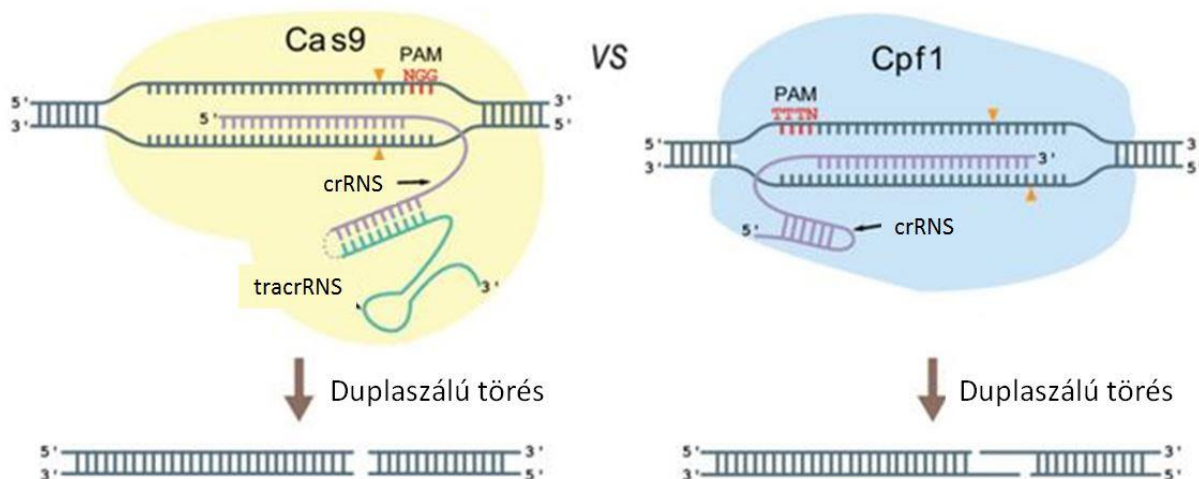


## 5. Ábra. A CRISPR/Cas off-target hatás csökkentésére irányuló törekvések a fehérje, vagy a gRNS módosításával

**a,** A felső panel a különféle megnövelt specificitású Cas9 változatokat ábrázolja a megjelenésük éve szerint időrendi sorrendben, az SpCas9<sup>2Pro</sup>-ig bezárólag. Alatta bal oldalon látható a fehérjeszerkezet modellezve, a Cas9 doménekkal. Mellette jobb oldalon a különböző megnövelt specificitású Cas9 változatok szekvenciájában létrehozott mutációk sematikus ábrázolása látható. **b,** A specificitás növelése érdekében alkalmazott gRNS módosítások vázlatos rajza. MP: 2'-O-metil-3'-foszfono-acetát. BNA<sup>NC</sup>: 2', 4'-BNA<sup>NC</sup> [N-Me]<sup>15</sup>. Ábra forrása: Zuo és mtsai, 2020 alapján.

### 2.3. Cas12a nukleázok

A közelmúltban Zetsche és mtsai megvizsgálták a Cas12a (másnéven Cpf1) nukleázokat, hogy alkalmazhatók-e génszerkesztési feladatokra<sup>35,37</sup>. Ez a munka a Cas12a-k olyan, a 6. ábrán is látható tulajdonságait tárta fel, amelyek általánosnak tűnnek ugyan a Cas12a nukleázok között, de a Cas9-ekre nem jellemzőek. A Cas12a egyetlen crRNS-el működik, ellentétben a Cas9-el, amely tracrRNS-t is igényel a target felismerő aktivitáshoz. A Cas12a-k által felismert PAM-motívum meglehetősen T-ben gazdag, ellentétben például az SpCas9-el, amely NGG szekvenciájú PAM-ot ismer fel. Míg a Cas9 nukleázok tompa DNS végeket hoznak létre a DNS kettős szál törésekor, a Cas12a 4 vagy 5 nukleotid hosszú 5' túlnyúlójú végeket eredményez<sup>35,37</sup>. Az úttörő tanulmányban Zetsche és mtsai 16 db Cas12a nukleázt vizsgáltak meg *in vitro*, melyből kettőt találtak alkalmasnak hatékony génszerkesztésre a HEK-293FT sejtekben. Az AsCas12a-t az *Acidaminococcus* sp. BV3L6, az LbCas12a-t a *Lachnospiraceae* ND2006 baktériumokból azonosították. Ezekben a kísérletekben kevés targeten és csak az NHEJ javítási útvonal kiaknázásával vizsgálták a Cas12a nukleázok aktivitását és eredményeik alapján az AsCas12a és LbCas12a által létrehozott indel gyakoriság magasabb volt, mint az SpCas9-é<sup>35,37</sup>.



#### 6. Ábra. A Cas9 és a Cas12a nukleázok összehasonlítása

Az ábra bal oldalán a Cas9-, jobb oldalon a Cas12a fehérje-RNS komplex látható, a fő működésbéli eltéréseikkel jelölve. A Cas9 tompa véget hoz létre a hasítás után, a Cas12a pedig ragadós véget. A Cas9 specificitását biztosító kis RNS két részből áll és az 5' vég felőli oldala tapad ki a DNS-re, a Cas12a kis RNS-e egy részből áll és a 3' vég felőli oldala tapad ki a DNS-re. Az SpCas9 az NGG PAM-ot, az Lb- és az AsCas12a a TTTN PAM-ot ismeri fel. Míg a Cas9-nél a DNS hasítása a PAM-hoz közel, a Cas12a-nél a PAM-tól távol történik<sup>38</sup>. Ábra forrása: Vanegas és mtsai, 2019 alapján.

Az előző fejezetben említésre került, hogy a Cas12a nukleázokról többen kimutatták, hogy alacsonyabb off-target-aktivitással bírnak, mint az SpCas9. A genom egészére kiterjedő off-target elemzések és a célzott mélyszekvenálás azt sugallják, hogy a Cas12a nukleázok csak egy vagy két bázis eltérést tolerálnak, ellentétben az SpCas9-el, amelynél arról számoltak be, hogy akár 5-6 nukleotidos eltérést is tolerálnak a gRNS spacer szekvenciája és a target szekvenciája között<sup>34-36</sup>.

#### **2.4. A CRISPR/Cas9 működését befolyásolja a gRNS-ek mennyisége és hossza**

A CRISPR rendszer hatékonyságát többféleképpen is befolyásolhatja a gRNS. Ezek közül az egyik legfontosabb, hogy megfelelő mennyiségű RNS legyen jelen, amely kötődni tud a nukleázhoz. Több csoport kimutatta, hogy a Cas9 működéséhez elengedhetetlen a guideRNS-ek megfelelő mennyiségű jelenléte. Különböző alkalmazásokban adott telítési szintig növelve a guideRNS-ek mennyiségét, növelhető a Cas9 hatékonysága, a nem megfelelő mennyiségű gRNS pedig akár teljesen eliminálhatja az aktivitást<sup>39-42</sup>. Emiatt kiemelkedően nagy szükség van arra, hogy a guideRNS expresszió megfelelő hatékonyságú legyen a CRISPR-es alkalmazásoknál.

Számos kutatás szerint a guideRNS-ek mennyiségén túl, azok 5' végének hossza is nagymértékben befolyásolja a Cas9 nukleáz aktivitását. Azt találták, hogy míg a guideRNS-ek 5' végének 1-2 bázissal való rövidítése még nem, addig 3 vagy több bázissal való csonkolása már rontja az SpCas9 nukleáz aktivitását, a 4 bázissal való rövidítés pedig már szinte teljesen eliminálja<sup>22,43,44</sup>. A guideRNS-ek 5' végének rövidítése egyaránt képes hatni az on- és off-target aktivitásra. Fu és mtsai azt találták, hogy az 5' végükön 1-2 nukleotiddal rövidebb guideRNS-ek off-target hatása kisebb, on-target aktivitása viszont hasonló mértékű, mint a normál hosszúságú gRNS-eké<sup>22</sup>. Tsai és mtsai később szintén megerősítették, hogy a 2-3 nukleotiddal rövidebb gRNS-ek alacsonyabb off-target aktivitással bírnak<sup>45</sup>.

A guideRNS-ek 5' végének azonban nemcsak a rövidítése, de a hosszabbítása is befolyásolhatja az SpCas9 off-target és on-target aktivitását. Többen azt találták, hogy egy plusz G nukleotid hozzáadása a guideRNS 5' végén csökkenti az SpCas9 off-target hatását. Wienert és mtsai azt találták, hogy közben az on-target hatékonyság hasonló volt az 5' G nélküli és a G-vel meghosszabbított guideRNS-ek használatakor<sup>46</sup>. Cho és mtsai viszont megmutatták, hogy a plusz G nukleotidok a guideRNS 5' végén nemcsak az off-target aktivitást csökkentik, de közben az on-target hatékonyság is csökkent<sup>47</sup>. A gRNS-ek expressziója általában az 5' végükön egy G nukleotiddal történik, melynek okait részletesebben a következő fejezetben mutatom be. Ezért vagy olyan targeteket választanak, amelyek guaninnal kezdődnek, vagy a

guideRNS-eket módosítják a spacerben, hogy a 20. (a 20. pozíció a 2. ábrán látható), vagy 21. pozícióban G legyen. Ez a guideRNS 5' végén helyezkedik el, ami viszont az előbbieken bemutatott eredmények szerint módosíthatja a nukleáz hasítási hatékonyságát. Míg azonban az SpCas9 akár a guideRNS-ek 5' végének 1-3 nukleotiddal való rövidítését, valamint az 5' vég plusz G-al való hosszabbítását is jól tolerálja, sőt ezek az off-target hatást is csökkenthetik, addig azok az SpCas9 variánsok, amelyeket úgy módosítottak, hogy magasabb specificitással bírjanak, jóval kevésbé tolerálják ezeket a módosításokat, és az on-target aktivitásuk jelentősen csökkenhet ilyen guideRNS-ek használatával<sup>48-50</sup>.

A guideRNS-eket kémiai is lehet szintetizáltatni, amellyel pontosan szabályozható az RNS hossz és mennyiség, viszont alkalmazásuk jelentős költségekkel jár. A gRNS-ek előállítására emiatt két költséghatékonyabb módszer terjedt el a gyakorlatban: a T7 promóter segítségével *in vitro* RNS transzkripció, illetve a különböző organizmusokban más-más kis RNS expresszióra alkalmas promóterek használata.

A fejezetben bemutatott eredmények alapján elmondható, hogy a CRISPR/Cas rendszerek használata során kiemelkedően nagy szükség van arra, hogy a guideRNS-ek hossza és mennyisége megfelelő legyen. Következésképp nagy igény van a gRNS expresszióra gyakran alkalmazott promóterek alapos megismerésére, amellyel fény derülhet arra, hogy melyik promóterrel hogyan érhető el a leghatékonyabb RNS expresszió és a megfelelő hosszúságú gRNS-ek legnagyobb aránya.

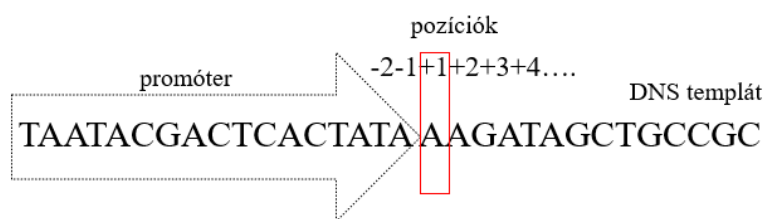
## **2.5. A CRISPR gRNS-ek expressziójára alkalmazott promóterek**

A CRISPR/Cas rendszert általában vagy plazmidokról expresszálni, vagy *in vitro* expresszálni és az organizmusba RNS-fehérje komplexet juttatva alkalmazták. Mind a nukleáz, mind a spaceret tartalmazó kis RNS-ek, akár az *in vitro*, akár a plazmidról történő expressziójához a legtöbb organizmusban néhány ismertebb promótert használnak. Emlős sejtekben a gRNS-ek expressziójára leggyakrabban alkalmazott promóterek az U6, a 7sk és a H1, *in vitro* a T7. *E. coli* baktériumokban pedig gyakran használt promóter a J23119.

Több tanulmány szerint a +1-es pozícióban elhelyezkedő nukleotid befolyásolhatja az RNS expressziót és a transzkripció iniciáció pozícióját, ezáltal az RNS-ek 5' végének hosszát is<sup>51-53</sup>. A promóter szekvenciájától downstream elhelyezkedő első bázist nevezik +1-es pozíciónak, várhatóan innen történik általában az első RNS bázis szintézise a DNS templát alapján.

Az ettől downstream elhelyezkedő pozíciókat a promótertől való távolságuk alapján számozzák és plusszal jelölik. A +1-től upstream lévő pozíciók számozása ugyanilyen módon történik és mínusszal jelölik őket (7. ábra).





## 7. Ábra. A +1 és a környezetében elhelyezkedő pozíciók számozása

A promóter szekvenciájától downstream elhelyezkedő első bázist +1-es pozíciónak nevezik. Tőle upstream mínusszal, downstream pedig plusszal jelölik a pozíciókat. Az ábrán a T7 promóter szekvenciája szerepel.

Az U6, 7sk, H1 promóterek a 3. típusú Polimeráz III (PolIII) promóterek közé tartoznak, melyek a Pol III promóterek között olyan szempontból egyedinek tekinthetők, hogy kizárólag upstream szabályozó elemeket tartalmaznak. Ide tartozik a proximális szekvencia elem (PSE), disztális szekvencia elem (DSE) és a TATA box (8. ábra). Általában jól meghatározott transzkripció kezdő pozíciójuk van, a +1-es pozíció, valamint egy polyT szekvencia a terminációs szignáljuk. Ennek köszönhetően ezek a promóterek, mint például a 7SK, U6 és a H1 szinte bármilyen kis RNS expressziójára könnyedén felhasználhatók. Ezen promóterek használata széles körben elterjedt kis RNS-ek expresszióra, többek között shRNS-ek expressziójához alkalmazzák őket RNS interferenciánál, valamint guideRNS expresszióra CRISPR/Cas rendszerekhez<sup>51</sup>.

Bár a H1 promótert is gyakran alkalmazzák, de többen megállapították, hogy az iniciációs pozíciója nem fix, ezért különböző hosszúságú RNS-ek íródnak át róla<sup>51,52</sup>.

Gao és mtsai azt találták, hogy az U6 és a 7sk humán promótereknél befolyásolja az átírás kezdő pozícióját a +1-es pozícióban elhelyezkedő nukleotid. Mindkét promóterrel akkor pontosan a +1 a transzkripció iniciációjának pozíciója, ha a +1-ben purin van<sup>51</sup>. Ma és mtsai a humán U6 promóterről 3 különböző, de a +1-ben G-t tartalmazó szekvenciát vizsgáltak meg új-generációs szekvenálással (NGS), néhány ezer readdal. Ők szintén azt találták, hogy ebben az esetben a transzkripció iniciációja majdnem mindig a +1-ről történik<sup>52</sup>. Többen azt találták, hogy az U6 és a 7sk promóternél nemcsak az átírás kezdő pozícióját, de az expressziós szintet is befolyásolja a +1-es pozícióban lévő nukleotid. Az elmúlt években leírták, hogy mindkét promóter +1A-val is hatékonyan használható guideRNS-ek expressziójára<sup>49</sup>. Gao és mtsainak eredményei szerint a +1-es pozícióban az A és a G nukleotid közel azonos RNS szintet biztosít mindkét promóternél<sup>51,54</sup>.



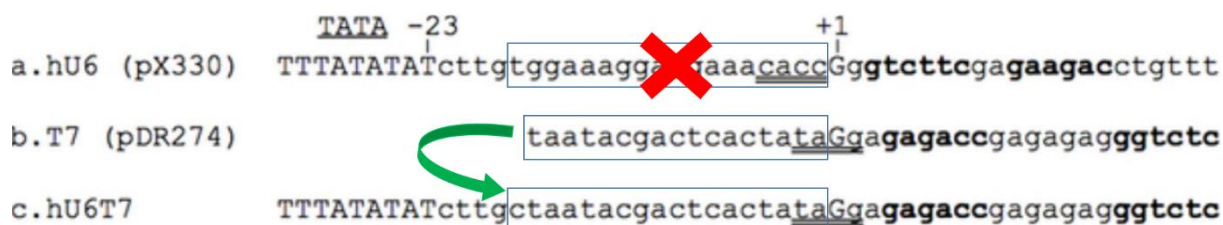
Ma és mtsai azt is megállapították, hogy olyan módosított humán U6 promótereknél, amelyeknél a transzkripció iniciáció pozíciójától upstream lévő nukleotidokat módosították restriktív endonukleáz hasítóhelyek bevitelével, még a +1-es pozícióban purinnal sem fix az iniciáció pozíciója és nemcsak a +1-ről indul el az átírás<sup>52</sup>.

Név		Organizmus	+1-ről eddigi eredmény
U6		humán	A és G: hasonló mennyiségű és pontos hosszúságú RNS-ek
7sk		humán	A és G: hasonló mennyiségű és pontos hosszúságú RNS-ek
U6T7		hibrid	G-vel mutatták be
T7		fág	G preferencia
J23119		<i>E.coli</i>	?

#### 8. Ábra. CRISPR gRNS expresszióra gyakran alkalmazott promóterek

A táblázatban a CRISPR gRNS-ek expressziójára alkalmazott promóterek közül látható néhány, pár releváns tulajdonságukkal. Középen egy nagyon sematikus ábrázolásuk látható néhány szekvencia elemükkel, úgy, mint a disztális szekvencia elem (DSE, sötétszürke box), proximális szekvencia elem (PSE, világosszürke box) és a TATA box (kék box). Az ábrán szerepel az is, hogy milyen organizmusból származnak, valamint összefoglalva a +1-es pozícióról eddig megjelent irodalmi adatok.

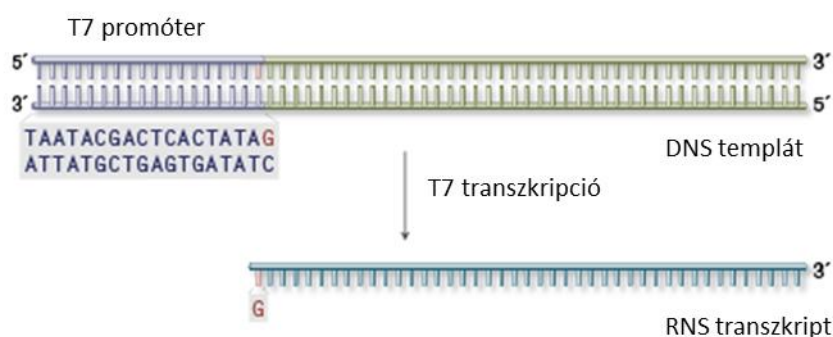
Romanienko és mtsai létrehoztak egy hibrid promótert, melyet U6T7-nek neveztek el, abból a célból, hogy ugyanaz a plazmid konstrukció alkalmas legyen emlős sejtekben és *in vitro* is kis RNS expresszióra. Az U6T7 hibrid promóter szekvenciájában abban különbözik az U6 szekvenciájától, hogy az U6 utolsó 18 bázispárnyi szekvenciát 3' irányból kicserélték a T7 promóter szekvenciájára, illetve egy bázist megváltoztattak (9. ábra). Megmutatták, hogy emlősökben és *in vitro* is hatékony expresszió érhető el vele. Az U6T7 promótert a promótertől downstream egy, vagy két G-vel javasolták használni, az eddigi U6 és T7 promóteres publikációk alapján<sup>55</sup>.



### 9. Ábra. Az U6T7 promóter létrehozása

Az ábrán a T7 promóter szekvenciája, valamint a humán U6 és az U6T7 hibrid promóterek szekvenciárészletei láthatók. Az U6T7 promótert úgy hozták létre, hogy a humán U6 promóter utolsó 18 bázispárnyi szekvenciáját 3' irányból kicserélték a T7 promóter szekvenciájára, illetve egy bázist megváltoztattak<sup>55</sup>. Ábra forrása: Romanienko és mtsai, 2016 alapján.

A T7 RNS polimeráz (T7 RNAP) egy rövid promóterrel rendkívül hatékony expresszióra képes. A T7 promótert széles körben használják fehérjetermeléshez baktériumsejtekben és nagy mennyiségű RNS előállítására *in vitro*, mint például aptamerek, ribozimek, siRNS-ek, CRISPR rendszer RNS-ei. Habár a legtöbbször a +1-es pozícióban guaninnal használják a T7 promótert, többen leírták, hogy alacsonyabb hatékonysággal *in vitro* +1 A-val is működik az expresszió, sőt kevésbé hatékonyan, de primidinről is megy az átírás<sup>53,56</sup>. Ezekben a korai kísérletekben, melyek nem a mai legkorszerűbb technikákkal készültek, megmutatták, hogy míg guaninnal általában a +1-ről indul az átírás, addig a többi esetben heterogénebb hosszúságú RNS-ek keletkeznek<sup>53</sup>. Az azonban a szakirodalomban nem konzisztens, hogy a T7 promótert egy darab G-vel (10. ábra), vagy 2<sup>47,55,57,58</sup>, vagy akár 3 G-vel érdemes használni<sup>58-62</sup>, és jelenlegi ismereteink szerint e 3 variáció esetében nem történt még olyan átfogó kutatás, amely szisztematikusan összehasonlította volna, hogy melyik variációról mennyi RNS expresszálódik, ezáltal melyik a legalkalmasabb variáció expresszióra.



### 10. Ábra. T7 promóter

Az ábrán a T7 promóter sematikus ábrázolása látható a szekvenciájával együtt. A promótert általában a promótertól downstream legalább egy G nukleotiddal használják, amely az ábrán is látható. A TATA szekvenciától downstream elhelyezkedő +1-es pozícióban G van, mely így a promóterrel átíródó RNS 5' első nukleotidja is. Ábra forrása: <https://international.neb.com/protocols/2015/11/24/sgna-synthesis-using-the-hiscribe-quick-t7-high-yield-rna-synthesis-kit-neb-e2050> oldal alapján.

A J23119 promótert CRISPR guideRNS-ek bakteriális expressziójára széles körben alkalmazzák *E. coli*-ban<sup>63-66</sup>. A J23119 egy mindössze 35 bp hosszú konstitutív promóter szekvencia és az Anderson promóter család legerősebb expresszióra képes tagja<sup>67</sup>. Eddig ismereteink szerint nem látott napvilágot olyan eredmény, amely megmutatta volna, hogy a J23119 promótertől downstream milyen kezdő nukleotidot, vagy nukleotidokat érdemes alkalmazni, úgy, mint ahogy a T7, vagy az U6 és 7sk promótereknél.

Ahogy fent bemutatásra került, a 7sk és az U6 promóterről már ismert, hogy a +1-es pozícióban A-val is használható. Ennek ellenére általában G-vel alkalmazzák őket<sup>49,51,54</sup>, a T7 promótert pedig minimum egy guaninnal használják<sup>47,55,57-62</sup>. Emiatt az SpCas9 gRNS-ekhez általában olyan targeteket választanak a gyakorlatban, amelyekben a spacer 5' első bázisának megfelelő bázis G, ezáltal a választható targetek köre szűkül. Ennek feloldására vagy a guideRNS-eket módosítják a 20., vagy 21.-ik pozícióban G-vel, ami viszont módosíthatja a nukleáz hasítási hatékonyságát<sup>46-50</sup> (lásd 2.4. fejezet). A promóterek vizsgálatával kapcsolatos eddigi kutatásokban csak a +1-es pozíciót vizsgálták szisztematikusan, azt is csak a T7, U6 és 7sk promótereknél. Az eddigi eredmények alapján feltételezhető, hogy az összes gRNS expresszióra alkalmazott promóternél befolyásolhatja a +1-es pozíció az expressziót és az RNS-ek hosszát és elképzelhető, hogy nemcsak a +1-es pozíció, hanem más pozíciók is.

Abból az okból kifolyólag, hogy a CRISPR/Cas rendszert befolyásolja a gRNS-ek mennyisége és hossza, valamint hasznos lenne a CRISPR-el használható targetek körét bővíteni, felmerül az igény a gRNS-ekre gyakran alkalmazott promóterek alaposabb megismerése.

### 3. Célkitűzések

Az elmúlt években Zetsche és mtsai megmutatták az As- és az LbCas12a nukleázokról HEK-293FT sejtekben, hogy alkalmasak génszerkesztésre<sup>37</sup>. Úttörő publikációjukat követően az elsők között kívántuk megvizsgálni ezen új nukleázok alkalmazhatóságát génmanipulációs eljárásokra más emlős sejtekben is. Az SpCas9 mellett kiválasztottunk három olyan Cas9-et, amelyek az SpCas9 után a leggyakrabban alkalmazott nukleázok közé tartoznak: a *Staphylococcus aureus*-ból (SaCas9)<sup>68</sup>, a *Streptococcus thermophilus*-ból (StCas9)<sup>69</sup> és a *Neisseria meningitidis*-ből (NmCas9)<sup>69</sup> azonosított nukleázokat és ezek aktivitását kívántuk összehasonlítani a Cas12a nukleázokkal. Az összehasonlításukhoz a korábban publikált GFxFP esszét<sup>70</sup>, illetve annak általunk továbbfejlesztett változatát kívántuk felhasználni. Vizsgálataink során a Cas12a nukleázokkal kapcsolatosan az alábbi kérdésekre kerestük a választ:

1. Alkalmas-e a továbbfejlesztett GFxFP riporter esszé a különböző nukleázok hasítási aktivitásának vizsgálatához?
2. Milyen hatékonysággal képesek az Lb-és AsCas12a nukleázok az SSA javítás indukálására néhány gyakrabban használt Cas9 nukleázhoz képest?
3. Alkalmasak-e az Lb-és AsCas12a nukleázok egy donormolekula HR javítás általi beépítésére a genomba?

A megnövelt specificitású SpCas9 nukleázok széles körű alkalmazhatósága nagymértékben függ attól, hogy képesek-e módosított gRNS-ekkel hatékonyan működni. A módosított gRNS-ek használata bevett gyakorlat szekvenciájukban nem G-vel kezdődő targeteknél. A megnövelt specificitású SpCas9 variánsokkal (melyek közül az eCas9-et és az SpCas9-HF1-et vizsgáltuk) kapcsolatosan az alábbi kérdésekre kerestük a választ:

1. Milyen hatással jár a 17, 18 és 19 nukleotid hosszúságú, 5' végükön rövidebb gRNS-ek alkalmazása a megnövelt specificitású SpCas9 variánsok aktivitására?
2. Hogyan befolyásolja az 5' végükön nem G-t, vagy egy nem komplementer G-t tartalmazó gRNS-ek alkalmazása a megnövelt specificitású SpCas9 variánsok aktivitását?

A CRISPR gRNS-ek expressziójára gyakran alkalmazott promóterek emlősökben az U6, 7sk, *in vitro* a T7, *E.coli*-ban a J23119 promóterek. Mivel a Cas nukleázok működését befolyásolja a gRNS-ek mennyisége és hossza, ezért rendkívül fontos, hogy pontos ismereteink legyenek az alkalmazott promóterekről. Több tanulmány szerint az, hogy milyen nukleotid követi a promótert, befolyásolhatja az RNS expressziót és a transzkripció iniciáció pozícióját is. Bár néhány általunk vizsgált promóter esetében a promótert követő első nukleotid szerepét már vizsgálták, de a további nukleotidok szerepét még nem. Ezek a +1-es pozícióra irányuló vizsgálatok ráadásul általában nem a jelenleg a kis RNS-ek hosszáról és mennyiségéről legpontosabb képet adó NGS módszerekkel történtek, hanem egyéb kevésbé pontos módszerekkel, illetve az eddig napvilágot látott kevés NGS-el kapott eredményhez nagyon kis számú readet alkalmaztak. A doktori értekezésem alapjául szolgáló kísérleteinkben a kiválasztott promóterek sokkal átfogóbb vizsgálatát kívántuk elvégezni az eddigiekhez képest. Ezt plazmidkönyvtárak létrehozásával és a róluk termelődött RNS-ek nagy readszámú újgenerációs szekvenálásával kívántuk megvalósítani, a promótert követő nemcsak az első, hanem az első 4 nukleotid szerepét megvizsgálva, ezáltal mélyrehatóbban tanulmányozva a megfigyelt promótereket. Doktori értekezésemben a promóterekkel kapcsolatosan az alábbi kérdésekre kerestük a választ:

1. Hogyan befolyásolja az U6 promótertől downstream lévő nukleotidok szekvenciája a transzkripció iniciációs pozíciót és az RNS expressziót?
2. Az U6 promóter esetében mennyire alkalmasak a +1A-val kezdődő szekvenciák kis RNS-ek expressziójára a +1G-val kezdődő szekvenciákhoz képest?
3. Hogyan befolyásolja a 7sk promótertől downstream lévő nukleotidok szekvenciája a transzkripció iniciációs pozíciót és az RNS expressziót?
4. Mennyi RNS expresszálódik a 7sk promóterről az U6 promóterhez képest?
5. Hogyan befolyásolja a T7 promótertől downstream lévő nukleotidok szekvenciája a transzkripció iniciációs pozíciót és az RNS expressziót?
6. A T7 promóterről egy, kettő, vagy 3 G-val expresszálódik több RNS?
7. Megváltoztatja-e az U6 promóter módosítása U6T7 hibrid promóterre a transzkripció iniciációs kezdő pozíciót?
8. Hogyan befolyásolja a J23119 promótertől downstream lévő nukleotidok szekvenciája a transzkripció iniciációs pozíciót és az RNS expressziót?

## 4. Anyagok és módszerek

### 4.1. Plazmidok létrehozása a Cas12a nukleázok és a megnövelt specificitású SpCas9 variációk vizsgálatához

A plazmidokat az inaktív nukleázok kivételével (ld. deadLbCas12a) standard molekuláris biológiai technikák alkalmazásával állítottuk elő. Az összes elkészült konstrukciót Sanger szekvenálással ellenőriztük (Microsynth AG). A disszertációban bemutatott eredményekhez az általam, vagy segítségemmel elkészített plazmid konstrukciók létrehozása részletesebben az alábbiak szerint történt.

#### *A pGF-chl-FP plazmid*

A pCAG-EGxxFP plazmidot (Addgene: 50716) EcoRV enzimmal emésztettem, melybe egy kloramfenikol rezisztencia gént ligáltam egy a csoportunkban korábban létrehozott plazmidból. A kloramfenikol kazettát PstI és BglII enzimekkel emésztettem, majd a végeket Klenow-fagmenttel töltöttem fel.

#### *A pGF-ori-FP plazmid*

A pCAG-EGxxFP plazmidot (Addgene: 50716) EcoRV enzimmal emésztettem, melybe a pUC replikációs ORI-t ligáltam egy általam korábban létrehozott plazmidból. A kazettát BclI és BglII enzimekkel emésztettem, majd a végeket Klenow-fagmenttel töltöttem fel. Ezután kiejtettem az eredeti ORI-t Eco31I és SacI enzimekkel emésztve a vektort.

#### *Spacer klónozások*

A vektorokat vagy Addgene-ről rendeltük, vagy a csoportunkban korábban készültek. Az SpCas9 spacer tartalmazó gRNS expressziós plazmidokat úgy állítottuk elő, hogy a spacer szekvenciát kódoló szintetizált DNS-oligonukleotidokat 4 nt hosszúságú túlnyúlóval egy BbsI restrikciós enzimmal emésztett pmCherry gRNS (Addgene: 80457) plazmidba ligáltuk. Egylépéses emésztés-ligálási protokollt követtünk. A szintetikus DNS oligonukleotidokat hibridizáltuk, és a duplaszálú linkereket (2,5  $\mu$ M) 50 ng plazmiddal, 3 egység BbsI restrikciós enzimmal és 1,5 egység T4 DNS ligázzal összekevertük 500  $\mu$ M ATP-t tartalmazó Green pufferben (Thermo Fisher Scientific). Az elegyet egy órán át 37 °C-on tartottuk, mielőtt kémiai kompetens, Stable *E. coli* sejtekbe (NEB) transzformáltuk. Az LB agar lemezen végzett tenyésztés után létrejött kolóniák közül kettőből a plazmidokat restrikciós enzimes emésztéssel teszteltük és a megfelelő klónokat elküldtük szekvenálni.

Az SaCas9 spacereket ugyanígy klónoztuk, de a vektort (Addgene: Px601) Eco31I enzimmel emésztettük, az St- és NmCas9 spacereket pedig Esp3I-vel. Az As-és LbCas12a plazmidokat szintén így állítottuk elő, és a vektort (Addgene: pTE4396, pTE4398) szintén Esp3I enzimmel emésztettük. Az Esp3I emésztésekhez 1 mM DTT-t (Sigma-Aldrich) és Tango puffert használtunk.

#### *dead LbCas12a-et kódoló plazmid*

Az inaktív LbCas12a-et úgy hoztam létre, hogy a 832-es aminosavat alaninra mutáltam. Az LbCas12a vektort (pY016, Addgene: 69988) két fragmensben PCR-el amplifikáltam, és a primerekben a két fragmensre egymással átfedő szakaszok voltak, az egyik primerpár pedig a D832A mutációt hordozta. A két PCR fragmenst *E.coli* kompetens sejtekbe transzformáltuk, homológ rekombinációs hibajavításon alapuló klónozási eljárással<sup>71</sup>.

A Cas12a nukleázok vizsgálataihoz felhasznált további plazmid konstrukciók az Irodalomjegyzékben a 35., a megnövelt specifitású SpCas9 variációkkal kapcsolatos vizsgálatokhoz pedig a 48. számmal jelzett publikációinkban megtalálhatók.

## **4.2. Sejttenyésztés**

Az N2a sejteket (Neuro-2a egér neuroblasztoma sejtek, ATCC - CCL-131) és az N2a sejtekből a csoportunkban létrehozott N2a.EGFP sejteket (olyan sejtvonal, amely az EGFP kazetta egyetlen kópiájú integrált példányát tartalmazza, amely a PRNP promóterről expresszálódik), valamint a HEK-293 (Gibco 293-H) sejteket 37 °C-on tenyésztettük párásított atmoszférában, 5% CO<sub>2</sub> tartalom mellett, magas glükóztartalmú Dulbecco's Modified Eagle's tápközegben (DMEM). A médiumot 10% hővel inaktivált magzati szarvasmarha szérummal, 4 mM L-glutammal (Gibco), 100 egység/ml penicillinnel és 100 µg/ml sztreptomocinnal egészítettük ki.

## **4.3. A GFxFP vizsgálat**

A GFxFP esszé működésének részletes bemutatása az 5.1.1-es fejezetben található.

A sejteket 48 lyukú lemezekre raktuk ki egy nappal a transzfekció előtt  $3 \times 10^4$  sejt/lyuk sejtszámmal. 24 órával később körülbelül 40% -os konfluencia mellett a sejteket a plazmid-konstrukciókkal transzfektáltuk Turbofect reagenssel (Thermo Fisher Scientific), röviden az alábbiak szerint: 250 ng plazmid DNS-t és 1 µl Turbofect-et összekevertünk 50 µl szérummentes DMEM-ben, és az elegyet 30 percig inkubáltuk szobahőmérsékleten,

majd hozzáadtuk a sejtekhez. A transzfekciós tápközeget a sejteken 24 órás inkubálás után friss DMEM-re cseréltük ki. Minden mintából három párhuzamos transzfektálást hajtottunk végre. Mindegyik mintával együtt mCherry expressziós plazmidot kotranszfektáltunk a transzfekciós hatékonyság ellenőrzése céljából. A sejteket áramlási citometriával elemeztük két nappal a transzfekció után.

#### **4.4. Genomi homológ rekombinációs teszt**

N2a sejtekben a PRNP gént célzó crRNS-t tartalmazó As- és LbCas12a nukleázokat kódoló plazmidokat kotranszfektáltuk egy donor plazmiddal, és megfigyeltük a GFP-t expresszáló sejtek számát. A donor molekula egy promóter nélküli GFP nyitott leolvasási keretet (ORF) tartalmazott, amelyet 1000 bp hosszúságú, a PRNP génnel homológ karok fognak közre. A célzott integráció után a PRNP ORF helyébe a GFP ORF kerül, ahol a GFP expresszióját a PRNP gén promótere vezérli.

A sejteket 6 lyukú platekre tettük ki egy nappal a transzfekció előtt  $1 \times 10^5$  sejt/lyuk sejtszámmal. Másnap, körülbelül 30% -os konfluencia mellett, a sejteket transzfektáltuk a plazmid-konstrukciókkal Turbofect reagenssel az alábbiak szerint: összesen 4000 ng plazmid DNS-t és 4  $\mu$ l Turbofect-et összekevertünk 400  $\mu$ l szérummentes DMEM-ben, és az elegyet szobahőmérsékleten 30 percig inkubáltuk mielőtt hozzáadtuk a sejtekhez. A sejteken lévő transzfekciós tápközeget 24 órás inkubálás után friss DMEM-re cseréltük ki. Mintánként három párhuzamos transzfektálást hajtottunk végre. Két nappal a transzfekció után a sejteket tripszinnel felszedtük és a következőképpen osztottuk tovább: a sejtek 10%-át új 6 lyukú lemezekre tettük, a sejtek 90%-át pedig áramlási citométer segítségével elemeztük. 9 nappal a transzfekció után a mintákat ismételten megmértük áramlási citométerrel.

#### **4.5. EGFP hasítási teszt**

Az esszé során az N2a.EGFP sejtek EGFP szekvenciáját céloztuk a vizsgált nukleázokkal és az aktivitásuk méréséhez az EGFP jel csökkenését figyeltük meg a kontrollokhöz képest.

Az N2a.EGFP sejteket egy nappal a transzfekció előtt 48 lyukú lemezekre raktuk ki körülbelül  $3 \times 10^5$  sejt/lyuk sűrűséggel. A transzfektálást Turbofect transzfekciós reagenssel végeztük a gyártó által ajánlott protokollnak megfelelően. A transzfekciókat három párhuzamossal hajtottuk végre. Az N2a.EGFP sejteket kétféle plazmiddal együtt transzfektáltuk: SpCas9 expressziós plazmiddal (137 ng), valamint gRNS-t és mCherry-t kódoló plazmiddal (97 ng), 1  $\mu$ l Turbofect reagenssel, lyukanként.



A transzfekció hatékonyságát az mCherry-t expresszáló sejtek mCherry szignálja alapján számoltuk ki, melyet 72 órával a transzfekció után mértünk le áramlási citométerrel. Az EGFP szignált ~ 72, és ~ 168 órával mértük meg a transzfektálás után. Az EGFP háttérszintjét minden kísérlethez a nem transzfektált sejtek alapján illetve kétféle kontroll felhasználásával határoztuk meg: (i) egy nem EGFP targettel komplementer spacer tartalmazó gRNS-t és egy mCherryt kódoló plazmid valamint egy aktív SpCas9 plazmid kotranszfekciójával; vagy (ii) egy deadSpCas9 expressziós plazmid és egy EGFP-t célzó gRNS-t és mCherry-t kódoló plazmid kotranszfekciójával.

Az EGFP hasítás százalékos arányát az alábbiak szerint számoltuk. Az egyes mintákban az EGFP-pozitív sejtek százalékos arányát levontuk a kontrollokból kapott EGFP pozitív százalékos arányának átlagából, és a transzfekciós hatékonyság értékével súlyoztuk. A transzfekciós hatékonyság értékét a mintákban levő mCherry felhasználásával kaptuk: megmértük az egyes mintákban az mCherry-pozitív sejtek százalékos arányát, és kiszámoltuk azok eltérését az összes transzfektált mintán egyesével az mCherry-pozitív sejtek átlagos százalékos arányától  $(\text{Átlag mCherry} - \text{Minta mCherry}) / (\text{Átlag mCherry})$ . Mindegyik mintánál három párhuzamost dolgoztunk fel és értéküket átlagoltuk.

#### **4.6. Áramlási citometria**

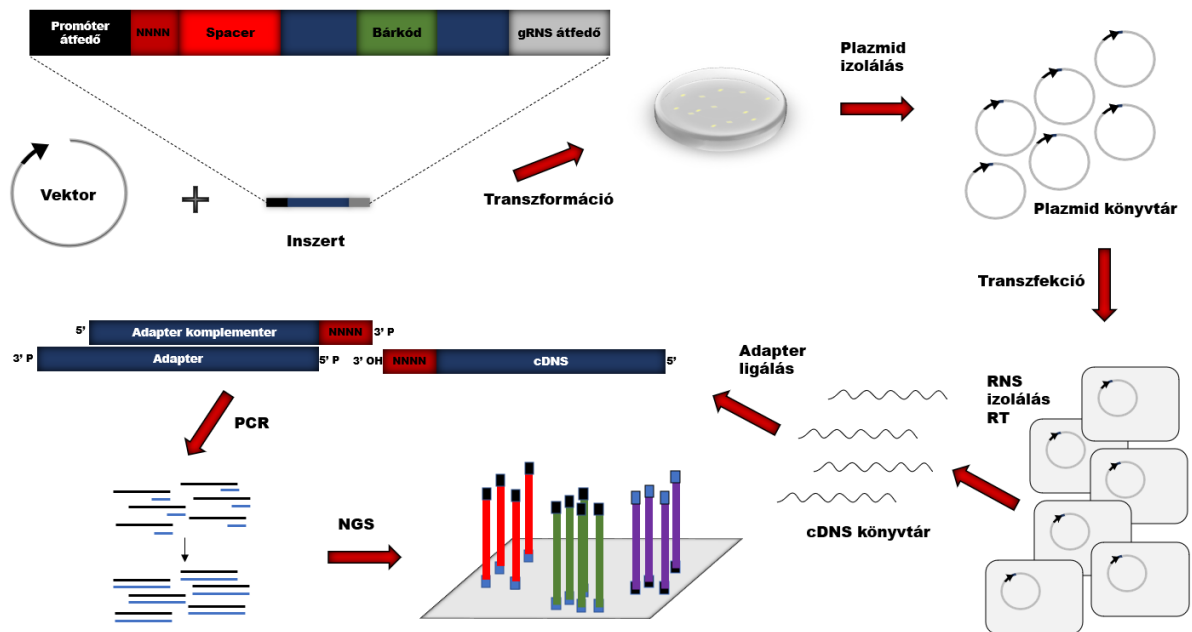
Az áramlási citometriát az Attune Acoustic Focusing citométerrel (Applied Biosystems) és a BD FACSCanto II-vel (Becton Dickinson Immunocytometry Systems), valamint a CytoFLEX Flow citométerrel (Beckman Coulter) végeztük. Az adatok elemzéséhez Attune Cytometric Software v.2.1.0 és CytExpert 2.0 verziókat használtuk. Az életképes egyes sejteket az oldalsó és a menetirányú fényszórási paraméterek alapján kapuztuk, és az összes kísérlet során összesen 5000–10 000 életképes egyedi sejtet vettünk figyelembe mintánként. Az Attune akusztikus fókuszáló citométer paraméterei: a GFP jel detektálásához 488 nm-es diódás lézert alkalmaztunk az excitációhoz és az emisszióhoz 530/30 nm-es szűrőt. Az mCherry jel detektálásához 488 nm-es diódás lézert alkalmaztunk az excitációhoz, 640LP szűrőt az emisszióhoz. A CytoFLEX áramlási citométer paraméterei: a GFP fluoreszcencia jelet 488 nm-es diódás lézerrel gerjesztettük és 525/40 nm-es szűrőt használtunk az emisszióhoz; az mCherry fluoreszcens jelet 638 nm-es diódás lézerrel gerjesztettük és a 660/20 nm-es szűrőt használtunk. A BD FACSCanto II áramlási citométer paraméterei: az iRFP670 jelének detektálásához 633 nm-es dióda lézert használtunk és 660/20 nm-es szűrőt.

#### 4.7. Plazmid könyvtárak létrehozása

A promóterektől downstream lévő nukleotidok szerepének vizsgálatához olyan plazmid könyvtárakat hoztunk létre, amelyekben a promótert követő első 4 nukleotid randomizált és az RNS-t kódoló régióban további 12 degenerált nukleotid van. Ez a 12 bázispáros bárkód szolgál arra, hogy a különböző hosszúságú és mennyiségű RNS-ek kiindulási mennyiségét, valamint a DNS templatjaik szekvenciáját azonosíthassuk.

Plazmid könyvtár készítés lépései (11. ábra):

1. Plazmid könyvtár vektorainak létrehozása
2. Plazmid könyvtár inszertek létrehozása
3. A vektorok és az inszertek összeillesztése
4. Transzformáció
5. Plazmid könyvtár tisztítása



#### 11. Ábra. Promóterek elemzéséhez alkalmazott módszer lépései

A folyamatábrán a promóterek elemzéséhez alkalmazott módszer lépései láthatóak sematikusan a plazmid könyvtár vektorainak és inszertjeinek összeillesztésétől kezdve az NGS-ig, az emlős promóteres könyvtárakra jellemzően. Először létrehoztuk a plazmid könyvtárakat a vektor és inszert összeillesztésével. Az inszertek két végének szekvenciái megegyeztek a linarizált vektorok végeinek szekvenciájával, azaz a promóter és a gRNS szekvenciáival. Az inszertekben a promótert követő első 4 nukleotid randomizált, valamint a gRNS-ek egy további 12 bázispáros randomizált szakaszt, bárkódot is tartalmaznak. Az összeillesztést a transzformáció követte, amely után a plazmid könyvtárak tisztítása következett. A plazmid könyvtárakat az emlős sejt promóterek esetében HEK-293 sejtekbe transzfektáltuk. Ezt követte az RNS-ek izolálása, majd a reverz transzkripció egy gRNS-re specifikus primerrel. A cDNS-ek 3' végére adaptereket ligáltunk egyedi módszerrel. Az adapter egy részben duplaszálú DNS molekula. Az adapter mindkét vége és az adapterrel komplementer oligó 3' vége foszforiláltak. A komplementer oligó 3' végén 4 randomizált nukleotid segíti elő a szintén randomizált nukleotidokat tartalmazó cDNS könyvtárakkal való ligálást. A ligátumokat PCR-el amplifikáltuk, majd NGS-el szekvenáltuk, a kiindulási plazmid könyvtárakkal együtt.

#### 4.7.1. Plazmid könyvtár vektorainak létrehozása

A könyvtárakhoz először létre kellett hozni azokat a rekombináns plazmidkonstrukciókat, amelyek a kiválasztott promótereket és az SpCas9 gRNS-t tartalmazzák. Egy PUC plazmid alapú vektor volt a háttér, amely ampicillin rezisztenciát és egy nagy kópiaszámú replikációs origót tartalmazott. Ebbe klónoztam be a különböző promótereket és az SpCas9 guideRNS-t a pX330-U6-Chimeric\_BB-CBh-hSpCas9 plazmidból (Addgene: 42230), egy hasítóhelyeket tartalmazó, így könnyedén cserélhető spacerrel. Az U6 promótert a pX330-U6-Chimeric\_BB-CBh-hSpCas9 plazmidból (Addgene: 42230), a 7sk promótert a ph7SK-gRNA plazmidból (Addgene: 53189) PCR-el amplifikáltam, a J23119 promótert és az U6T7 promóterhez a T7 promóter szekvenciáját pedig linker klónozással építettem be a PUC háttérű plazmidokba. Az összes elkészült konstrukciót Sanger szekvenálással ellenőriztük (Microsynth AG). A plazmid könyvtárak vektorainak létrehozása a fent ismertetett módon előállított plazmidokból PCR-es amplifikálással történt. A PCR-ekhez Q5 polimeráz (NEB) enzimet használtunk, majd DpnI (ThermoFischer Scientific) emésztést követően a vektor PCR-t NucleoSpin Gel and PCR Clean-up (Macherey-Nagel) kittel tisztítottuk. A könyvtár vektor amplifikátumaihoz használt primereket az 1. Táblázat tartalmazza.

Plazmid könyvtár inszertjeinek oligói		
Név	Forward primer	Reverz primer
U6 spori	TATCTTGTGGAAAGGACGAAACACC	TTTTCAAGTTGATAACGGACTAGCCTTATTTAACTTGCTA
U6 prp	TATCTTGTGGAAAGGACGAAACACC	TTTTCAAGTTGATAACGGACTAGCCTTATTTAACTTGCTA
U6 1551	TATCTTGTGGAAAGGACGAAACACC	TTTTCAAGTTGATAACGGACTAGCCTTATTTAACTTGCTA
7sk spori	AGCTTGTGCGCCGCTTGG	CCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTAACTTGCTA
7sk prp	AGCTTGTGCGCCGCTTGG	CCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTAACTTGCTA
7sk 1551	AGCTTGTGCGCCGCTTGG	CCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTAACTTGCTA
U6T7 spori	CTTGCTAATACGACTCACTATA	CCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTAACTTGCTA
U6T7 prp	CTTGCTAATACGACTCACTATA	CCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTAACTTGCTA
U6T7 1551	CTTGCTAATACGACTCACTATA	CCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTAACTTGCTA
BBaj23119 spori	GTCTAGGTATAATGCTAGC	CCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTAACTTGCTA
BBaj23119 prp	GTCTAGGTATAATGCTAGC	CCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTAACTTGCTA
BBaj23119 1551	GTCTAGGTATAATGCTAGC	CCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTAACTTGCTA

Plazmid könyvtár vektorainak oligói		
Név	Forward primer	Reverz primer
U6 spori	GGTGTTTCGTCTTTCCACAAGATA	TAGCAAGTTAAAATAAGGCTAGTCCGTTATCA
U6 prp	GGTGTTTCGTCTTTCCACAAGATA	TAGCAAGTTAAAATAAGGCTAGTCCGTTATCA
U6 1551	GGTGTTTCGTCTTTCCACAAGATA	TAGCAAGTTAAAATAAGGCTAGTCCGTTATCA
7sk spori	GCGCACAAGCTATATAAACCTGAAGGAAGTCT	CTTGAAAAAGTGGCACCAGTCTG
7sk prp	GCGCACAAGCTATATAAACCTGAAGGAAGTCT	CTTGAAAAAGTGGCACCAGTCTG
7sk 1551	GCGCACAAGCTATATAAACCTGAAGGAAGTCT	CTTGAAAAAGTGGCACCAGTCTG
U6T7 spori	TATAGTGAGTCGTATTAGCAAGATATATAAGCC	CTTGAAAAAGTGGCACCAGTCTG
U6T7 prp	TATAGTGAGTCGTATTAGCAAGATATATAAGCC	CTTGAAAAAGTGGCACCAGTCTG
U6T7 1551	TATAGTGAGTCGTATTAGCAAGATATATAAGCC	CTTGAAAAAGTGGCACCAGTCTG
BBaj23119 spori	GCTAGCATTATACCTAGG	TAGCAAGTTAAAATAAGGCTAGTCCGTTATCA
BBaj23119 prp	GCTAGCATTATACCTAGG	TAGCAAGTTAAAATAAGGCTAGTCCGTTATCA
BBaj23119 1551	GCTAGCATTATACCTAGG	TAGCAAGTTAAAATAAGGCTAGTCCGTTATCA

#### 1. Táblázat. A plazmid könyvtárak inszertjeihez és vektoraihoz szintetizált primerek szekvenciái

A plazmid könyvtárak inszert és vektor fragmenseit PCR-ekkel amplifikáltuk összeillesztésük előtt, a táblázat az ehhez felhasznált primerek szekvenciáit tartalmazza 5'-3' irányban.

#### 4.7.2. Plazmid könyvtár inszertek létrehozása

Az inszertek tartalmazták az SpCas9 gRNS szekvenciájának olyan módosított szakaszát, ahol a promóter utáni első 4 nukleotid a spacerben és egy további 12 bázispár hosszú, bárkódként használt szekvencia randomizált nukleotidokból áll. Az inszertekhez DNS oligókat szintetizáltattunk, melyek két végükön a vektor plazmidok két végével (promóter, illetve a gRNS) azonos átfedő szekvenciákat tartalmaztak. Az inszert oligók sematikus ábrázolása a 11. ábrán látható, amely összefoglalja a promóterek vizsgálatához alkalmazott módszer további lépéseit is. Az inszerteket, melyekhez a szintetizált oligonukleotidok szekvenciái a 2. táblázatban találhatók, a szintetizált oligókból, PCR-el amplifikálva hoztuk létre, a primerek szekvenciái az 1. táblázatban vannak. A PCR-termékeket gélből izoláltuk, majd NucleoSpin Gel and PCR Clean-up (Macherey-Nagel) kittel tisztítottuk. A PCR-ekhez Q5 polimeráz (NEB) enzimet használtunk.

Név	Szekvencia 5'-3'
U6 spori	GGACTAGCCTTATTTAACTTGCTANNNNNNNNNNTAGCTCTAAACTGACACGTTCTTCGGNNNNNGGTGTTTCGTCCTTTCCACAAGATA
U6 Prp	GGACTAGCCTTATTTAACTTGCTANNNNNNNNNNTAGCTCTAAACCCCTCATCCCACGATCNNNNNGGTGTTTCGTCCTTTCCACAAGATA
U6 1551	GGACTAGCCTTATTTAACTTGCTANNNNNNNNNNTAGCTCTAAACGCTTGTCGCCAGGATNNNNNGGTGTTTCGTCCTTTCCACAAGATA
7sk spori	GGACTAGCCTTATTTAACTTGCTANNNNNNNNNNTAGCTCTAAACTGACACGTTCTTCGGNNNNNGAGGTACCCAAGCGGCGCACAAGCT
7sk Prp	GGACTAGCCTTATTTAACTTGCTANNNNNNNNNNTAGCTCTAAACCCCTCATCCCACGATCNNNNNGAGGTACCCAAGCGGCGCACAAGCT
7sk 1551	GGACTAGCCTTATTTAACTTGCTANNNNNNNNNNTAGCTCTAAACGCTTGTCGCCAGGATNNNNNGAGGTACCCAAGCGGCGCACAAGCT
U6T7 spori	GGACTAGCCTTATTTAACTTGCTANNNNNNNNNNTAGCTCTAAACTGACACGTTCTTCGGNNNNNTAGTGAGTCGTATTAATCGCC
U6T7 Prp	GGACTAGCCTTATTTAACTTGCTANNNNNNNNNNTAGCTCTAAACCCCTCATCCCACGATCNNNNNTAGTGAGTCGTATTAATCGCC
U6T7 1551	GGACTAGCCTTATTTAACTTGCTANNNNNNNNNNTAGCTCTAAACGCTTGTCGCCAGGATNNNNNTAGTGAGTCGTATTAATCGCC
BBaj23119 spori	GGACTAGCCTTATTTAACTTGCTANNNNNNNNNNTAGCTCTAAACTGACACGTTCTTCGGNNNNAGTATTATACCTAGGACTGAGC
BBaj23119 Prp	GGACTAGCCTTATTTAACTTGCTANNNNNNNNNNTAGCTCTAAACCCCTCATCCCACGATCNNNNAGTATTATACCTAGGACTGAGC
BBaj23119 1551	GGACTAGCCTTATTTAACTTGCTANNNNNNNNNNTAGCTCTAAACGCTTGTCGCCAGGATNNNNAGTATTATACCTAGGACTGAGC

#### 2. Táblázat. A plazmid könyvtár inszertjeihez szintetizált oligonukleotidok szekvenciái

A plazmid könyvtárak inszertjeihez oligonukleotidokat szintetizáltattunk, ezek szekvenciái láthatóak a táblázatban 5'-3' irányban.

#### 4.7.3. A vektorok és inszertek összeillesztése

Az inszertek és a vektorok összeillesztését Gibson klónozási technikával végeztük<sup>72</sup>, amely NEBuilder HiFi DNA Assembly Master Mix-el (NEB) történt, a protokoll szerint, 50 °C-on 1 órán át. Az inszert:vektor moláris aránya 3:1 volt.

#### 4.7.4. Transzformáció

A plazmid könyvtárakat ultrakompetens Dh5alpha (NEB) sejtekbe mértem, majd jégen 15 percig inkubáltam. Ezután hűsokkoltam 42 °C-on 1 percig, amit 400 ml LB rámerése után 15 perc rázás követett 37 °C -on. Az inkubáció után kiszélesztettem a könyvtárakat és egy éjszakán át 37 °C -on inkubáltam.

#### **4.7.5. Plazmid könyvtárak tisztítása**

A könyvtárak telepszámainak becslése után LB-vel mostam le a lemezeket, melyekből ezután GenElute HP Plasmid Miniprep (Sigma-Aldrich) kittel tisztítottam ki a plazmid könyvtárakat. Az összes könyvtárat Sanger szekvenálással ellenőriztük (Microsynth AG).

#### **4.8. A cDNS könyvtárak létrehozása**

A cDNS könyvtár készítés lépései (11. ábra):

4.8.1. Emlős sejtek RNS könyvtárainak létrehozása

4.8.2. Bakteriális promóteres RNS könyvtárak létrehozása

4.8.3. *In vitro* T7 transzkripció

4.8.4. Reverz transzkripció

##### **4.8.1. Emlős sejtek RNS könyvtárainak létrehozása**

Az U6, 7sk és U6T7 emlős promóteres plazmid könyvtárakat HEK-293 sejtekbe transzfektáltuk. A sejtenyésztésről további információ a 4.2. fejezetben található. A sejteket T75 méretű flasksokban szaporítottam fel, a kiindulási sejtszám 4 millió/flaska volt és a sejtek a kirakás után 24 órával kerültek transzfektálásra. A transzfekciós ágens Jetfect reagens (Biospirál-2006. Kft.) volt, egy flaskára 15 µg plazmidkeverék került, amely a plazmidkönyvtárból és egy kontroll GFP-t expresszáló plazmidból állt 9:1 arányban. Transzfekció után 48 órával a transzfekciós hatékonyságot a sejtek 10%-ából citometriásan az Attune Acoustic Focusing citométerrel, az Attune Cytometric szoftverrel ellenőriztük (lásd 4.6. fejezet). A sejtek kirakása után 72 órával a sejtek 90%-ából RNazol (Molecular Research Center) segítségével tisztítottuk az RNS-t. Az RNS-ek integritását minden mintánál gélelektroforézissel ellenőriztük. Minden plazmid könyvtárról 3 párhuzamos RNS minta készült. Az emlős sejtis promóterek vizsgálatához alkalmazott módszer lépései sematikusan a 11. ábrán láthatók.

##### **4.8.2. Bakteriális promóteres RNS könyvtárak létrehozása**

A bakteriális promóteres könyvtárakat tartalmazó Dh5alpha sejtekből 30%-os glicerines stockot készítettünk, melyet -20 °C- on tároltunk. A stockból 100 µl-t 10 ml ampicillint tartalmazó LB médiumban egy éjszakán át 37 °C -on rázatva inkubáltam. Az RNS-eket másnap RNazol (Molecular Research Center) segítségével izoláltam, a sejtszám körülbelül  $1 \times 10^9$  volt. Az RNS-ek integritását minden mintánál gélelektroforézis segítségével ellenőriztem. Minden plazmid könyvtárról 3 párhuzamos RNS minta készült.

#### 4.8.3. *In vitro* T7 transzkripció

Az U6T7 plazmid könyvtárakról a 3. táblázatban megadott primerekkel amplifikáltuk a T7 promótert és a guideRNS-t. Az amplifikátumokat gélből izoláltuk, majd NucleoSpin Gel and PCR Clean-up (Macherey-Nagel) kittel tisztítottuk. A templát PCR-ekről *in vitro* TranscriptAid T7 High Yield Transcription Kit segítségével expresszáltam az RNS-eket. Az RNS-eket az RNA Clean & Concentrator (Zymo) RNS tisztító kittel izoláltam. Az RNS-ek integritását minden mintánál gélelektroforézissel ellenőriztük. Minden plazmid könyvtárról 3 párhuzamos RNS minta készült.

Név	Szekvencia 5'-3'
T7 Forward	CTTGCTAATACGACTCACTATA
T7 Reverz	AAAAAAGCACCGACTCG

#### 3. Táblázat. Az *in vitro* T7 RNS transzkripció templátjához felhasznált primerek szekvenciái

A táblázatban a T7 promótereket és a gRNS-eket tartalmazó kazetta PCR-es amplifikálásához felhasznált primerek láthatók 5'-3' irányban. A primerekkel amplifikált PCR-ek szolgáltak templátul az *in vitro* transzkripcióhoz.

#### 4.8.4. Reverz transzkripció

A reverz transzkripcióhoz Maxima H minus first strand cDNA synthesis kit with dsDNase (Thermo Scientific) enzimet használtuk. A reverz transzkripció primer a következő volt: 5' GCCACTTTTCAAGTTGATAACGGAC 3'. A reakció 65 °C-on 1 órán át zajlott.

#### 4.9. Adapterek ligálása a cDNS-re

A cDNS 3' végére az adaptert T4 DNS ligáz (Thermo Fischer Scientific) segítségével ligáltuk. Az adapter egy részben duplaszálú DNS molekula volt, melyet a ligálás előtt hibridizáltunk. Az adapter 5' és 3' vége és az adapterrel komplementer oligó 3' vége foszforilációval kerültek szintetizálásra a Sigma-Aldrich által. A ligálás 37 °C-on, 1 órán át történt, 20%-os PEG8000 (Thermo Fischer Scientific) koncentráció mellett, T4 DNS ligáz (Thermo Fischer Scientific) enzim segítségével (lásd részletesebben a 18. ábrán és az 5.3.2. fejezetben).

#### 4.10. NGS

A plazmid könyvtárakat és a cDNS könyvtárakat Illumina Nextera túlnyúlókat tartalmazó specifikus primerekkel amplifikáltuk, majd a termékeket az NGS indexekkel meghosszabbított primerekkel amplifikáltuk egy következő PCR-el. Az első körös PCR primerek megtalálhatók a 4. táblázatban. A 2. PCR után a minták koncentrációit Qubit dsDNA HS (Thermo Fischer Scientific) kit segítségével határoztuk meg, és elvégeztük a megfelelő koncentrációkban történő összekeverést (pooling) az NGS-hez. A PCR-ket a BGI szekvenálta HiSeq X10-el, pair end technológiával. A szekvenálás minőségét a FastQC v.0.11.8 segítségével ellenőriztük. Az adapterek szekvenciáit Trimmomatic 0.38-mal trimmeltük a readokról. A forward és reverz readeket a FLASH-1.2.11 programmal párosítottuk.

#### 4.11. Programozás

Az előző pontban leírt lépéseket követően a további adatfeldolgozást Python-ban végeztük, mely az alábbi főbb lépésekből állt:

1. A gRNS szekvenciáját tartalmazó readek indexek szerinti szétválogatása
2. A kapott readek illesztése a gRNS szekvenciájára
3. Bárkód szekvenciák dinamikus felismerése (readen belül változó pozícióban van a bárkód)
4. 4N szekvenciák dinamikus felismerése (readen belül változó pozícióban van a 4N)
5. RNS és plazmid könyvtárak készítése (4N+bárkód+sokaság)
6. Reads szám alapján szűrés
7. RNS könyvtárak normalizálása a DNS könyvtárral
8. RNS könyvtárak 4N és különböző hosszúság szerinti szétválogatása és összeszámolása

A további kiértékelést Microsoft Excel segítségével végeztük el.

#### 4.12. Statisztika

A promóterek és a megnövelt specificitású SpCas9 variációk vizsgálatához a statisztika a következőképp történt. Egyszempontos varianciaanalízist használtunk, melyhez Levene-próbával ellenőriztük a varianciakülönbségeket. Szignifikáns Levene-próba esetén a Welch varianciaanalízist használtuk a klasszikus ANOVA helyett. A minták különbségeink összehasonlításához ANOVA alkalmazásakor Tukey, szignifikáns Levene próba után pedig Games-Howell post-hoc tesztekkel végeztünk el. A promóterek vizsgálatához IBM SPSS v25 programmal készítettük el a statisztikát, a megnövelt specificitású SpCas9 variációk tesztelésével kapcsolatos statisztika pedig R 3.4.1-ben készült.



		Forward primer
Plazmid könyvtárak	Név	
	U6 spori	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTATCTTGTGGAAAGGACGAAACACC
	U6 prp	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTATCTTGTGGAAAGGACGAAACACC
	U6 1551	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTATCTTGTGGAAAGGACGAAACACC
	7sk spori	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAGCTTGTGCGCCGCT
	7sk prp	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAGCTTGTGCGCCGCT
	7sk 1551	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAGCTTGTGCGCCGCT
	U6T7 spori	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGCTTTATATATCTTGCTAATACGACTCAC
	U6T7 prp	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGCTTTATATATCTTGCTAATACGACTCAC
	U6T7 1551	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGCTTTATATATCTTGCTAATACGACTCAC
	T7 spori	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGATATATCTTGCTAATACGACTCACTATA
	T7 prp	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGATATATCTTGCTAATACGACTCACTATA
	T7 1551	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGATATATCTTGCTAATACGACTCACTATA
	BBaj23119 spori	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGACCTAGGACTGAGCTAGCTGTCAA
	BBaj23119 prp	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGACCTAGGACTGAGCTAGCTGTCAA
	BBaj23119 1551	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGACCTAGGACTGAGCTAGCTGTCAA
cDNS könyvtárak	Összes adapter+cDNS	TCGTCGGCAGCGTCAGAT
		Reverz primer
Plazmid könyvtárak	U6 spori	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGCACCGACTCGGTGCCAC
	U6 prp	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGCACCGACTCGGTGCCAC
	U6 1551	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGCACCGACTCGGTGCCAC
	7sk spori	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGCACCGACTCGGTGCCAC
	7sk prp	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGCACCGACTCGGTGCCAC
	7sk 1551	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGCACCGACTCGGTGCCAC
	U6T7 spori	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGCACCGACTCGGTGCCAC
	U6T7 prp	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGCACCGACTCGGTGCCAC
	U6T7 1551	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGCACCGACTCGGTGCCAC
	T7 spori	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGCACCGACTCGGTGCCAC
	T7 prp	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGCACCGACTCGGTGCCAC
	T7 1551	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGCACCGACTCGGTGCCAC
	BBaj23119 spori	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGCACCGACTCGGTGCCAC
	BBaj23119 prp	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGCACCGACTCGGTGCCAC
	BBaj23119 1551	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGCACCGACTCGGTGCCAC
cDNS könyvtárak	Összes adapter+cDNS	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGGACTAGCCTATTTTAACTTGCTA

#### 4. Táblázat. A könyvtárak első NGS PCR primerei

A plazmid könyvtárakat és az adapterrel ligált cDNS könyvtárakat Illumina Nextera túlnyúlókat tartalmazó specifikus primerekkel amplifikáltuk, melyek a táblázatban láthatók 5'-3' irányban.



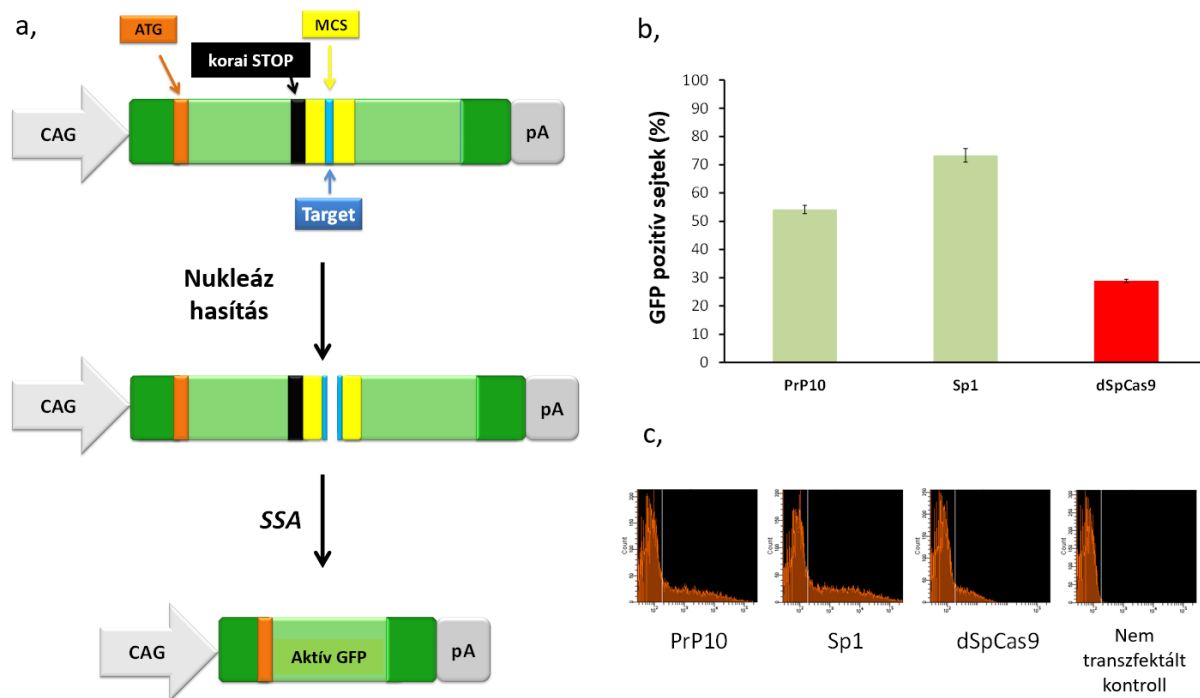
## 5. Eredmények

### 5.1. Cas12a nukleázok vizsgálata

#### 5.1.1. A GFxFP esszé módosítása

A fejezet összes eredménye, valamint ezekkel kapcsolatos további eredmények az Irodalomjegyzékben a 35. számmal hivatkozott publikációnkban található.

Az AsCas12a és LbCas12a nukleázok aktivitásának vizsgálatához a Mashiko és mtsai által publikált fluoreszcens riporter assay-t használtuk<sup>70</sup>, amely a 12. ábra a, részén látható. A GFP riporter plazmidban egy GFP-szekvencia 5' és 3' fragmentumát elválasztja egymástól két 480 bázispár hosszúságú ismétlődő szekvencia, amely a GFP szekvencia középső részlete. Az expressziós kazettát egy CAG promóter hajtja és egy SV40 (Simian Vacuolating Virus 40) polyA jel zárja le. Az első GFP fél egy STOP kodonnal fejeződik be, ezáltal erről a szakasztól nem észlelhető GFP fluoreszcencia. A második GFP félnek nincs sem Kozak szekvenciája, sem ATG kodonja. A nukleáz targetét egy poliklónozó helyre (MCS) kell klónozni a két GFP fél között. A nukleázok target szekvenciáját a két, ismétlődő szekvencia közé helyezve a nukleáz hasításakor létrehozott kettős szálú DNS-törés a homológ rekombinációs hibajavítás egyik altípusával, a single strand annealing DNS javító mechanizmussal (SSA) javítódik (ld. 2.1. fejezet). Ennek az esszének, amelyet „GFxFP assay-nek” hívunk az az előnye például egy genomi pozíciókat célzó esszével szemben, hogy bármilyen nukleáz DNS hasítási hatékonyságának mérését lehetővé teszi egy specifikus target hasításával, miközben a hasítás hatékonyságát nem befolyásolják külső tényezők. Ilyenek a genomi környezet, a target szekvencia epigenetikus állapota, vagy a target relatív távolsága a homológ karoktól.



## 12. Ábra. GFxFP riporter esszé

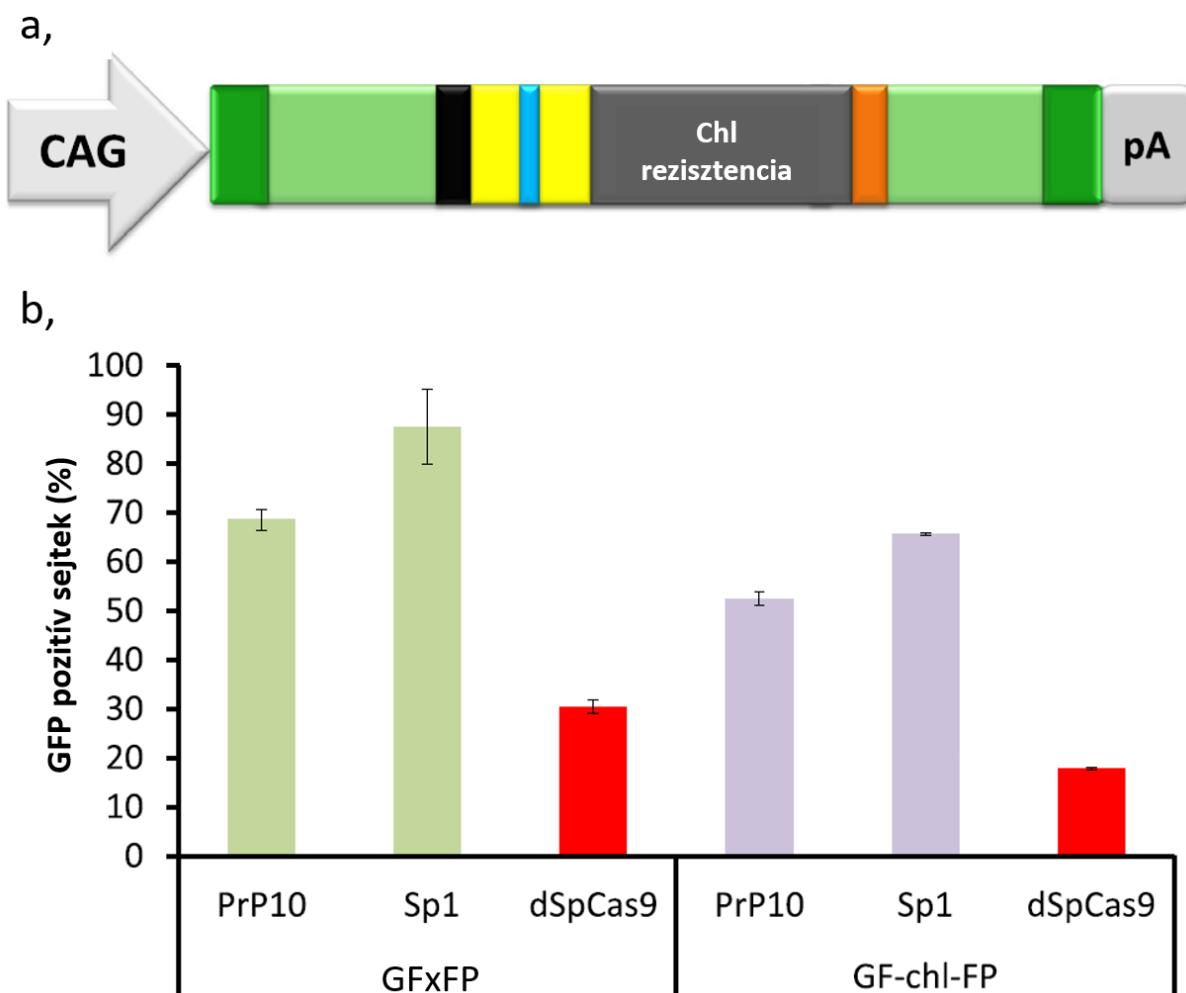
**a,** Az ábra a nukleázok aktivitásának megfigyelésére alkalmazott riporter esszé elvét ábrázolja. A két GFP-félszekvencia zölden van jelölve. Ezek mindegyike tartalmaz egy egyedi szegmenst (sötétzöld) és egy átfedő, azonos szekvenciát tartalmazó szegmenst (világoszöld). Az expressziós kazettát egy CAG promóter hajtja (szürke nyíl), és egy SV40 (Simian Vacuolating Virus 40) polyA jel (pA, szürke doboz) zárja le. Az első GFP fél egy STOP kodonnal fejeződik be (fekete jel). A nukleáz targetjét (kék téglalap) egy poliklonózó helyre (MCS, sárga) kell klónozni a két GFP fél között. A nukleáz hasításával a célszekvenciában létrehozott kettős szálú DNS-törést az ismétlődő szekvenciák által irányított SSA javítás javítja. Ennek eredményeként egy ép GFP szekvencia jön létre, amelyből egy aktív GFP íródik át. **b,** A GFP fluoreszcenciát mutató sejtek százalékos aránya. A Prp10 és Sp1 oszlopok azt mutatják meg, hogy hány %-ban tartalmaztak GFP-t azok a sejtpopulációk, amelyek a Prp10, vagy Sp1 targeteket tartalmazó GFxFP plazmidokkal és a megfelelő gRNS-eket és Cas9-et kódoló plazmidokkal kerültek transzfektálásra. A dSpCas9 olyan sejteket takar, amelyeket inaktív SpCas9-et expresszáló vektorral transzfektáltunk. Az értékek kiszámításához iRFP670 szignált használtunk a transzekciós hatékonyság mérésére és a GFP értékeket erre normalizáltuk. A hibásávok az  $n=3$  független transzekció során mért százalékok szórását mutatják be. **c,** A GFP fluoreszcencia intenzitások reprezentatív hisztogramjai, amelyeket a Prp10, Sp1, dSpCas9 és egy nem transzfektált sejtpopuláció mintáinak áramlási citometriás elemzésével mértünk a transzekció utáni második napon<sup>35</sup>.

Ennek a vizsgálati rendszernek a teszteléséhez SpCas9-et használtunk két target szekvenciával (PrP10- és Sp1-ként fogok rájuk utalni a következőkben). Az Sp1 vagy PrP10 targetet tartalmazó GFxFP plazmidokat N2a egér neuroblasztóma sejtekbe kotranszfektáltuk, SpCas9-et, a megfelelő guideRNS (gRNS)-t és iRFP670 fluoreszcens fehérjét expresszáló plazmidokkal. Kontrollként inaktív dSpCas9-et expresszáló vektort használtunk a sejtek transzfektálására. A fluoreszcens sejtek számát két nappal a transzekció után mértük le. Az iRFP670 a transzekció hatékonyságának nyomon követését segítette elő, míg a kontroll háttér feletti GFP szint növekedés mérése a hasítási hatékonyság vizsgálatát.

Az Sp1-target 73% -ot, míg a PrP10-target 54% -os fluoreszcens sejtszámot eredményezett a transzfektált populációkban (12. ábra b).

Meglepő módon a kontroll populáció is figyelemre méltó, körülbelül 30% -os GFP-pozitivitást mutatott. Látszólag a plazmidok egy részében a homológ szekvenciák közötti rekombináció a nukleáz hasítása nélkül is megtörtént. Ezután, annak érdekében, hogy megfigyelhessük a potenciális rekombinációs eseményeket, amelyek már a plazmid kezelése során megtörténhetnek a baktériumsejtekben, vagy kloramfenikol rezisztencia gént (pGF-chl-FP, 13. ábra a), vagy a plazmid replikációs origóját (pGF-ori-FP) klónoztam be a GFP fragmensek homológ szekvenciái közötti régióba. A baktériumsejtekből végzett plazmid izolálás után egyik esetben sem észleltünk rekombinált frakciót (pGF-ori-FP és pGF-chl-FP plazmidok), amikor restrikciós enzimes emésztéseket követően vizsgáltuk őket agaróz gélelektroforézissel. Ez arra utal, hogy egy ilyen populáció, ha létezik is, kevesebb, mint 1%.

Ezután ezekbe a vektorokba is beklónoztuk a korábban használt Sp1 és PrP10 target szekvenciákat, és megvizsgáltuk az SpCas9 által indukált GFP javítást N2a sejtekben. A kloramfenikol kazettát tartalmazó GFxFP plazmiddal az aktív SpCas9-tartalmú mintákban a fluoreszcens sejtek százalékos aránya kevesebb volt, mint a kloramfenikol nélkül alkalmazott GFxFP-vel. Amikor az inaktív dSpCas9-et expresszáló vektort pGF-chl-FP plazmiddal kotranszfektáltuk, a GFP-pozitív sejtek továbbra is kimutathatók voltak, bár alacsonyabb százalékban, mint a módosítatlan eredeti plazmid, a GFxFP esetében (13. ábra b). Ez arra utal, hogy a funkcionális GFP-molekulák még a pGF-chl-FP-vel módosított plazmidok egy részében is expresszázódtak a nukleáz hasítása nélkül is. A nukleáz hasítása mellett bekövetkező GFP jel csökkenése nemcsak a baktériumsejtekben már rekombinált plazmidok arányának csökkenésével magyarázható, hanem a két GFP fél csökkent rekombinációs képességét is tükrözi, mivel a köztük lévő távolságot növeli a kloramfenikol rezisztencia kazetta. A viszonylag magas háttér-GFP-fluoreszcencia eredete nem teljesen tisztázott, de azt feltételeztük, hogy az N2a-sejtekben természetesen előforduló kettős szálú törések kissé magas szintjéből származhat. Ennek kiderítésére anti-H2AX festést végeztünk többféle sejten, amellyel azt kaptuk, hogy az N2a és a HEK-293T sejtek összehasonlítható, de figyelemre méltóan magasabb szintű DNS kettős szál-törést mutatnak, mint a HeLa sejtek. Hasonló eredményeket találtunk a pGF-ori-FP plazmiddal is, majd ezt a plazmidot használtuk a további kísérletekben.



**13. Ábra. A kloramfenikol rezisztencia gén integrációja csökkenti a fluoreszcens sejtek számát a GFxFP vizsgálatban**

a, A GFxFP kazetta sematikus ábrázolása, a 12. ábrához hasonló módon.

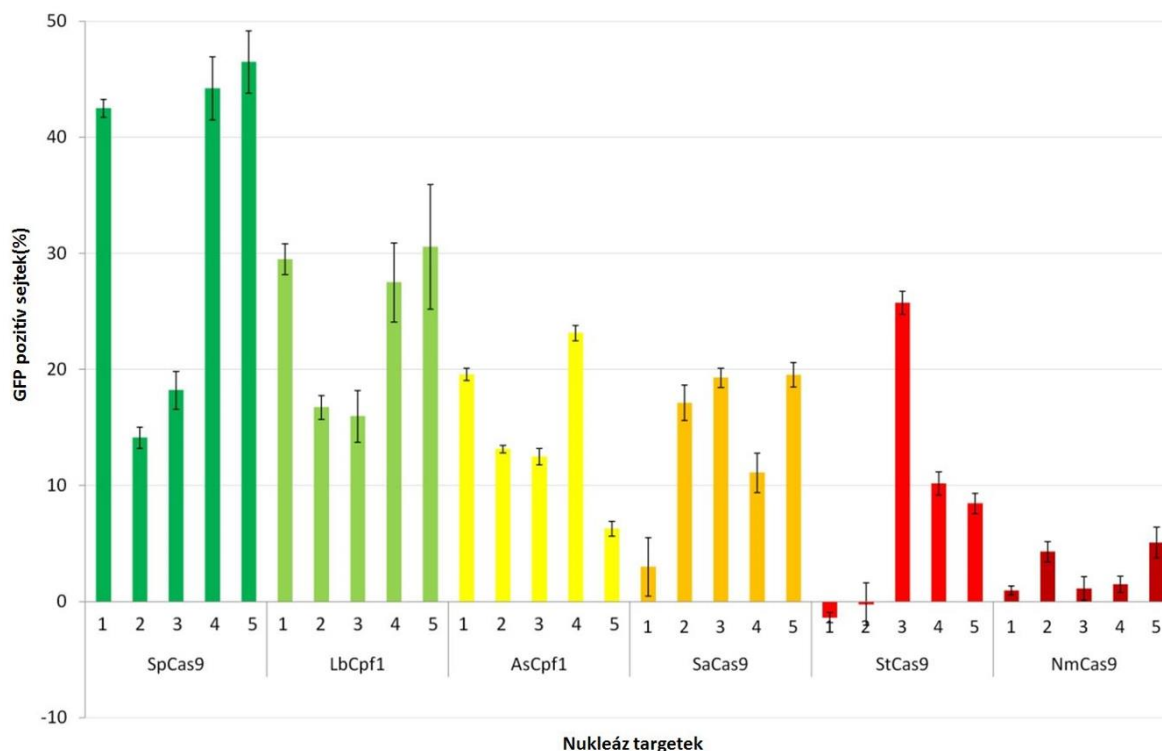
A kloramfenikol rezisztencia gén egy sötét szürke doboz közepén a GFP-fél fragmensek homológ szekvenciái között. b, A GFP pozitív sejtek százalékos arányának összehasonlítása, ha a riporter esszé az eredeti vagy (GFxFP, zöld oszlopok) vagy a kloramfenikol rezisztencia génnel kiegészített (GF-chl-FP, lila sávok) GFxFP plazmidokat tartalmaz. Mindegyik esetben két targetet teszteltünk, a PrP10-et és az Sp1-et, az SpCas9 nukleáz és a megfelelő gRNS-eket kódoló plazmidok kotranszfekeciójával N2a sejtekben. Kontrollként inaktív SpCas9-et használtunk mindkét esetben (piros oszlopok, dSpCas9). Az értékeket a transzfekeciós hatékonysággal normalizáltuk, iRFP670 fluoreszcencia alapján, amelyet transzfekeciós kontrollként használtunk. A hibaoszlopok a százalékos értékek átlagainak szórását mutatják. N = 3 volt. A kloramfenikol kazetta integrációja csökkentette a fluoreszcens sejtek számát mind az aktív, mind az inaktív SpCas9-el<sup>35</sup>.

### 5.1.2. Az As- és LbCas12a hatékonyan indukálja a GFP kazetta javítását a GFxFP esszében

Az As- és LbCas12a nukleázok hatékonyságának tesztelésére emlős sejtekben HR javítást indukáltunk 20 genomi target szekvenciát kiválasztva. Ezek három egér prionfehérje család gént céloztak meg: PRNP, SPRN és PRND. A targeteket a GFxFP plazmidba, a megfelelő spacereket az AsCas12a- és LbCas12a-crRNS-eket is kódoló plazmidokba klónoztuk.

A tesztelés során mind az As-, mind az LbCas12a hatékonyan indukálta a GFP kazetta javítását mind a 20 targettel, minden esetben több, mint 10% fluoreszcens sejtet eredményezve a kontrollhoz képest, megerősítve ezen Cas12a nukleázok aktivitásának robusztusságát (nincs ábrázolva).

Számos Cas9 nukleázt jellemeztek annak érdekében, hogy a legszélesebb körben alkalmazott SpCas9-hez hasonlóan használható nukleázt találjanak. Ezekből kiválasztottunk hármat, amelyek az SpCas9 után a leggyakrabban alkalmazott nukleázok közé tartoznak: a *Staphylococcus aureus*-ból (SaCas9)<sup>68</sup>, a *Streptococcus thermophilus*-ból (StCas9)<sup>69</sup> és a *Neisseria meningitidis*-ből (NmCas9)<sup>69</sup> azonosított Cas9 nukleázokat, és összehasonlítottuk robusztusságukat az AsCas12a és LbCas12a nukleázokkal, 5-5 különböző és nukleázonként eltérő targetek felhasználásával. Azért alkalmaztunk különböző targeteket, mert létező genomi pozíciókat kívántuk megvizsgálni, amelynek eredményeit későbbi kísérletekben is fel kívántuk használni és a targeteket a PAM szekvenciák alapján választottuk ki. A targeteket véletlenszerűen választottuk ki, azaz anélkül, hogy a nukleázok szekvencia specifitására vonatkozó előzetes ismereteket alkalmaztunk volna. Az SpCas9 esetében ez nem így történt, itt kihasználtuk a gRNS-hatékonyság-jósló eszközök létezését: sgRNA designer-<sup>73</sup> és az sgRNA scorer-t 1.0<sup>74</sup>, és az általuk közepes, vagy annál jobbnak ítélt targeteket választottuk ki. A többi nukleázhoz ekkor még nem léteztek ilyen predikciós programok. Az összes targetet a GFxFP riporter rendszerrel teszteltük, amely lehetővé teszi a nukleázok aktivitásának közvetlenebb összehasonlítását kizárva azt, hogy a targetek homológ karoktól lévő távolsága, vagy a genomi környezet befolyásolja a hasítás hatékonyságát. Az összes mintát mCherry expressziós vektorral együtt transzfektáltuk, és az eredményeket az mCherry pozitív populáción belül elemeztük. A negatív kontrollok GFP fluoreszcenciáját inaktív SpCas9-et expresszáló vektor alkalmazásával mértük, és a kapott értékeket kivontuk minden egyes minta értékéből. Minden esetben három párhuzamos transzfektálást hajtottunk végre. Eredményeink alapján ebben a vizsgálatban mindkét Cas12a nukleáz ugyanannyira, vagy hatékonyabban indukálta az SSA hibajavítást átlagosan, mint három vizsgált Cas9 (14. ábra). Az LbCas12a az 5 különböző target alapján átlagosan 24% az AsCas12a 15%-os GFP szint növekedést eredményezett a kontrollhoz képest, szemben a SaCas9 13%, StCas9 9% és az NmCas9 3% -os eredményéhez képest. A várakozásoknak megfelelően az SpCas9 mutatta a legerősebb aktivitást, bár ehhez hozzájárulhatott a predikciós programok használata is. Ezek az eredmények arra utalnak, hogy a Cas12a nukleázok fontos helyet foglalnak el az SpCas9 mellett a genomszerkesztő eszközök palettáján.



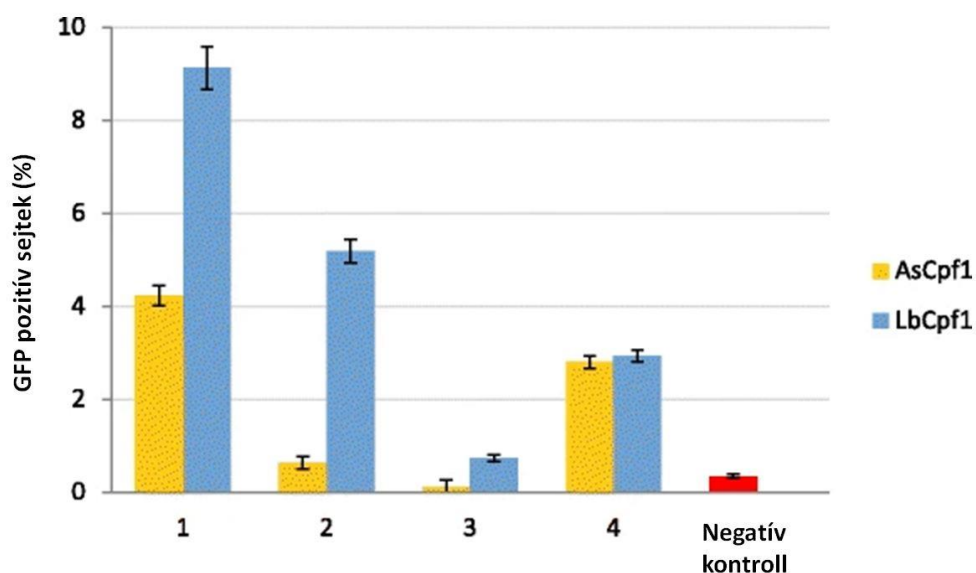
#### 14. Ábra. A különböző nukleázok aktivitásának összehasonlítása GFxFP-vel

A GFP háttér feletti GFP pozitív sejtek százalékos aránya, különféle nukleázok hatására N2a sejtekben. Hat nukleázot teszteltünk, mindegyiket öt különböző targeten, a GFxFP esszé alkalmazásával. Az oszlopok az egyes nukleázok esetében az öt különböző target hasítási hatékonyságára kapott átlagértékeket mutatják. A GFP pozitív sejteket a transzfekció után két nappal mértük. Az összes mintát mCherry expressziós vektorral együtt transzfektáltuk, és az eredményeket az mCherry pozitív populáción belül elemeztük. A negatív kontrollok GFP fluoreszcenciáját inaktív SpCas9-et expresszáló vektor alkalmazásával mértük, és a kapott értékeket kivontuk minden egyes minta értékéből. Minden targettel három párhuzamos transzfektálást hajtottunk végre. Az átlagos HR-indukáló hatékonyságot 15 minta átlagából számoltuk nukleázonként. A hibásávok a százalékos értékek átlagának szórását mutatják. SpCas9: zöld, LbCas12a: világos zöld, AsCas12a: sárga, SaCas9: narancs, StCas9: piros, NmCas9: skarlát<sup>35</sup>.

#### 5.1.3. Az As- és LbCas12a nukleázok által végzett genomi DNS-hasítás hatékony HR javítást képes indukálni

Következő lépésként leteszteltük, hogy az As- és LbCas12a nukleázok képesek-e HR javítást indukálni a genomban. Ehhez leteszteltünk négy targetet a genomi környezetükben egy donor molekula segítségével. A donor molekula egy promóter nélküli GFP nyitott leolvasási keretet (ORF) tartalmaz, amelyet 1000 bp hosszúságú, a PRNP génnel homológ karok fognak közre. A célzott integráció után a PRNP ORF helyébe a GFP ORF kerül, ahol a GFP expresszióját a PRNP gén promótere vezérli. A PRNP-t célzó crRNS-t tartalmazó As- és LbCas12a nukleázokat kódoló plazmidokat kotranszfektáltuk a donor plazmiddal, és megfigyeltük a GFP-t expresszáló sejtek számát. Meglepő módon a promóter nélküli donor plazmid az inaktív dSpCas9-et expresszáló plazmiddal kotranszfektálva is eredményezett némi GFP fluoreszcenciát, amikor a transzfekció után két nappal mértük a fluoreszcens jelet.

Ezt a tranziens fluoreszcens szignált a további kísérletekben arra használtuk, hogy összehasonlítsuk a különböző minták transzfekciós hatékonyságát. Ez a tranziens GFP jel körülbelül 9 napon belül lecsengett, ami már lehetővé tette az integrált GFP kazettákból származó GFP jel mérését. A 15. ábrán látható eredmény azt mutatja, hogy mindkét nukleáz képes elősegíteni a HR javítást, ezáltal a donor kazetta homológ integrációját az N2a sejtekbe. A négy vizsgált PRNP-célpont közül az LbCas12a hármat, míg az AsCas12a két targetet hasított, amelyek az inaktív nukleázzal kotranszfektált negatív kontroll szintjénél nagyobb arányban eredményeztek fluoreszcens sejteket. Hasonlóan alacsony fluoreszcenciaszintet találtunk egy olyan kontrollban is, ahol a Cas12a nukleázokat megfelelő homológ kar nélküli donor plazmiddal aktív nukleázzal kotranszfektáltuk (nincs ábrázolva), ami azt mutatja, hogy a Cas12a aktivitása nem eredményez szignifikáns fluoreszcenciát a nem HR javítás által közvetített donor véletlenszerű integrációjával. Ez valószínűleg annak köszönhető, hogy a GFP kazetta előtt nem található promóter a plazmidban, így az integrációja nem eredményez expressziót, hacsak nem pont egy promóter után integrálódik a genomba. A fejezet összes eredménye az Irodalomjegyzékben 35. számmal hivatkozott publikációnkból származik.



**15. Ábra. Az As- és LbCas12a-indukált HR javítás szintje genomi targeteken**

Az ábrán a GFP pozitív sejtek százalékos aránya látható egy promóter nélküli donor GFP kazetta HR javítás által közvetített genomi integrációja után. Mind az As-, mind az LbCas12a nukleázok (narancssárga és kék) hatékonyságát négy egér PRNP genomi célszekvencián teszteltük. A nukleázvektort és a homológ rekombinációs donor plazmidot kotranszfektáltuk N2a sejtekbe. Negatív kontrollként a sejteket a donor molekulával és egy inaktív SpCas9 expressziós plazmiddal kotranszfektáltuk. A transzfekció utáni hetedik napon lemértük a GFP pozitív sejteket. Minden mintához három párhuzamos transzfektálást hajtottunk végre. A hibasávok a százalékos értékek átlagainak szórását mutatják<sup>35</sup>.

## 5.2. Megnövelt specificitású SpCas9 nukleázok vizsgálata

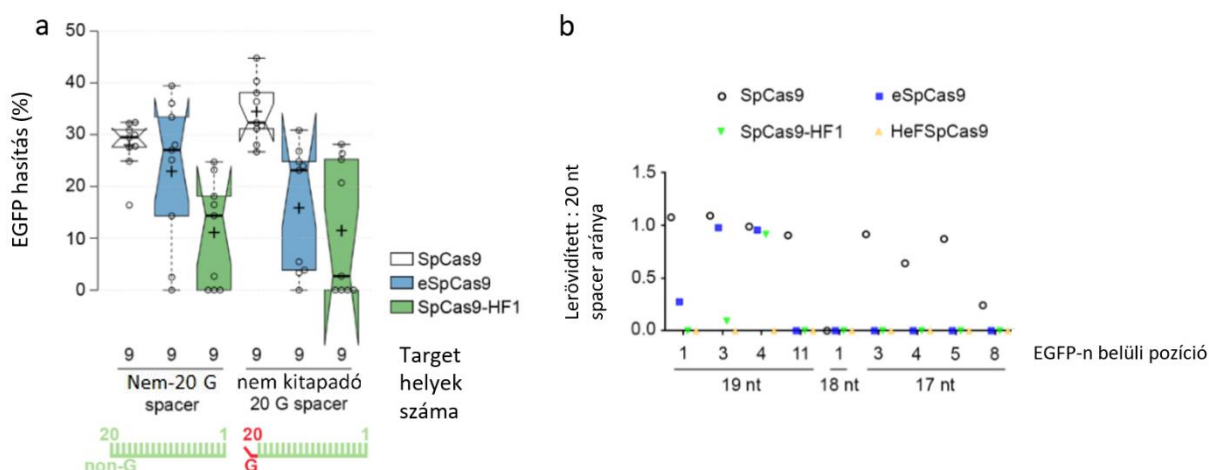
A fejezet összes eredménye, valamint ezekkel kapcsolatos további eredmények az Irodalomjegyzékben a 48. számmal hivatkozott publikációkban található.

Munkám ezen részében azt kívántuk megvizsgálni, hogy a megnövelt specificitású eSpCas9 és SpCas9-HF1 nukleázok aktivitását hogyan befolyásolja a módosított gRNS-ek alkalmazása. Ezen nukleázok széles körű alkalmazhatósága attól függ, hogy képesek-e a módosított gRNS-ekkel hatékonyan működni. Olyan targetek esetében, ahol a spacer 20. pozíciójának megfelelő nukleotid nem G, különféle gyakorlatok vannak a gRNS spacer 5'-végének módosítására (lásd 2.4. és 2.5. fejezet). A disszertációmban bemutatott munka során arra irányuló vizsgálatokat végeztünk, amelyekben a spacer 5' végének rövidítését, 5' végükön G-t nem tartalmazó gRNS-ek, vagy 5' végükön a 20. pozícióban egy nem komplementer G nukleotidot tartalmazó gRNS-ek hatását vizsgáltuk az SpCas9 variációk aktivitására. Utóbbi kettőt azért teszteltük, mert ezeknek a módszereknek a segítségével nemcsak G kezdetű targetekkel lennének használhatók a vizsgált megnövelt specificitású nukleázok. Az 5' végükön rövidített gRNS-eket azért teszteltük le, mert több cikk szerint ha az U6 promótert nem követi G, 5' végükön rövidebb RNS-ek keletkezhetnek<sup>75</sup>. Ahhoz, hogy a megnövelt specificitású SpCas9 nukleáz variánsok aktivitását megfigyeljük, számos targeten direkt összehasonlításokat végeztünk el EGFP hasítási teszttel. Ennek során N2a.EGFP sejteket használtunk. Ez egy olyan, a csoportunkban létrehozott sejtvonal, amely az EGFP kazetta egyetlen kópiájú integrált példányát tartalmazza Neuro-2a egér neuroblasztoma sejtekben, ami a PRNP promóterről expresszálódik. A nukleázokkal különböző EGFP szekvencián belüli targeteket céloztunk és az aktivitásuk szintjét az EGFP szignál csökkenésével mértük. A kísérletekben az N2a.EGFP sejteket kétféle plazmiddal kotranszfectáltuk: a különböző SpCas9 variációkat kódoló expressziós plazmiddal, valamint a különféle EGFP targeteknek megfelelő spaceret magukban foglaló gRNS-t és mCherry-t kódoló plazmiddal, és a sejteket áramlási citometriával értékeltük. Az mCherry fluoreszcenciája a transzfectió hatékonyság, az EGFP fluoreszcencia csökkenése pedig a nukleázok hasítási hatékonyságának meghatározására szolgált.



### 5.2.1. 5' végükön nem G-t, vagy a 20. pozícióban nem komplementer G-t tartalmazó gRNS-ek alkalmazása

Azoknál a targeteknél, amelyeknél a spacer 20. pozíciójának megfelelő nukleotid nem G, bevett gyakorlat a target szekvencia változtatás nélküli felhasználása, vagy a 20. pozíció G nukleotidra cserélése, amely így a targethez nem kitapadó 20. G-t eredményez a gRNS-ben. 18 gRNS-t teszteltünk le, melyeknek a fele egy 5' kitapadó nem-G nukleotidot tartalmaz, a másik felében pedig ugyanaz a spacer, de az 5' első nukleotid egy nem kitapadó G. Azt találtuk, hogy a WTSpCas9 valamivel jobban működik egy nem kitapadó kezdő G-vel, mint a nem G kezdetű spacerekkel (16. ábra a). Bár az eSpCas9 viszonylag jól működik a nem G kezdetű spacerekkel, és a WTSpCas9 aktivitásához mérten csak mintegy 20%-os csökkenés figyelhető meg a targetek többségénél (6/9), a 20. pozícióban a nem kitapadó G-re történő csere jelentősen csökkenti az aktivitását. Az SpCas9-HF1 lényegesen csökkent aktivitást mutat mindkét változatú módosított gRNS-el a WTSpCas9-hez képest.



**16. Ábra.** A gRNS különféle módosításainak hatása az SpCas9 nukleáz variánsok EGFP hasításának aktivitására

Az ábrán a spacer 5' végének módosításának hatása látható a nukleázok aktivitására. **a)** 20 nt hosszú 5' végén kitapadó, de nem G-t tartalmazó, és 20 nt hosszú 5' végén nem kitapadó G-t tartalmazó spacer hatása a nukleáz aktivitásra. **b,** A gRNS-ek 5' végének rövidítésének hatása a vad típusú SpCas9, eSpCas9, SpCas9-HF1 és HeFSpCas9 aktivitására. Az ábrán a rövidített gRNS-ekkel alkalmazott SpCas9-ek aktivitásának aránya van ábrázolva a normál hosszúságú gRNS-ekkel alkalmazott SpCas9-ek aktivitásával. Tukey-típusú box-plot a BoxPlotR által: a középvonalak a mediánokat mutatják; a boxok határai a 25. és 75. percentilist jelzik; a szimpla vízszintes vonal az interkvartilis tartomány 1,5-szereséig terjed ki a 25. és 75. percentilistől; a bevágások a boxokon a mediánok 95% -os konfidencia intervallumát jelzik; a keresztek a minta átlagát képviselik; az adatpontok körökként vannak ábrázolva. Az egyes nukleázok mintapontjának számai megfelelnek az alkalmazott targetek számának. Az adott kísérletekhez használt spacer típusai zöld színű fésűként vannak ábrázolva, és a 20. G nukleotidok (a PAM-szekvenciától számozva) piros színű, hajlított fésűfogakként vannak jelölve, ha nem tapadnak ki; a zöldek a kitapadó nukleotidokat képviselik<sup>48</sup>.

### **5.2.2. 5' végükön rövidebb gRNS-ek alkalmazása**

Annak vizsgálatára, hogy az SpCas9 variánsok miként tolerálják a rövidebb spacereket, 5 EGFP targetet választottunk ki, amelyből 9 db rövidített, 17, 18 vagy 19 nukleotid spacer hosszúságú gRNS-eket kódoló plazmidokat hoztunk létre, amelyek mindegyike kitapadó G-vel kezdődik. Ezekkel a gRNS-ekkel kotranszfektált SpCas9-ek EGFP hasítási aktivitását hasonlítottuk össze normál hosszúságú gRNS-ekkel kotranszfektált SpCas9-ek aktivitásával (16. ábra b). A kilenc csonkolt hosszúságú spacert tartalmazó gRNS-ek közül 8-al sikeresen hasította az EGFP-t a WTSpCas9. Ezzel szemben az eSpCas9 és az SpCas9-HF1 csak három, illetve egy targetet hasított, és ezek mind csak egy nukleotiddal rövidített, 19 nt hosszúságú spacerek voltak. Az általunk létrehozott, mind az eCas9, mind az SpCas9-HF1 mutációinak kombinációit tartalmazó HEFCas9 variáns egyik rövidített spacert tartalmazó gRNS-el sem működött.

Ezek a kísérletek azt mutatják, hogy a megnövelt specificitású Cas9 variánsok aktivitását csökkentik ezek a gRNS módosítások. A bemutatott változtatások az összes általunk vizsgált módosított Cas9 nukleáz aktivitását csökkentették, különböző mértékben. Az eSpCas9 alacsonyabb érzékenységet mutatott a módosításokra az SpCas9-HF1-hez képest.

A fejezet összes eredménye az Irodalomjegyzékben 48. számmal hivatkozott publikációnkból származik.

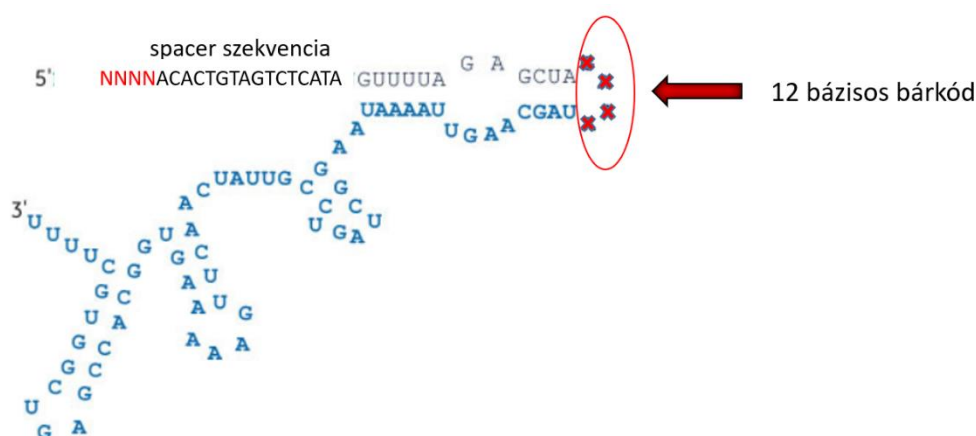
### 5.3. A CRISPR gRNS-ek expressziójára gyakran alkalmazott promóterek

A fejezetben szereplő eredményekből készülő kézirat jelenleg beküldés előtt áll.

#### 5.3.1. A randomizált bázisokkal módosított gRNS

A promóterek vizsgálatához olyan plazmid és ezekről létrehozott cDNS könyvtárakat készítettünk el, amelyekben a promóterektől downstream elhelyezkedő első 4 nukleotid randomizált volt. A transzkriptumok 3' irányban egy módosított SpCas9 gRNS-el folytatódtak. Az SpCas9 gRNS-t azért választottuk a promóterek vizsgálatához az RNS-t kódoló szekvenciaként, mert erős másodlagos szerkezete van a spacer-tól 3' irányban elhelyezkedő szakaszán. Ebből fakadóan azt képzeltük, hogy jobban tolerálja a randomizált szakaszokon belüli szekvencia különbségeket és a 12 bázisos bárkódot egy olyan helyre illesztettük, amiről azt felételeztük, hogy kevésbé változtatja meg a szerkezetét.

A 17. ábrán látható, hogy milyen másodlagos szerkezet jellemző az SpCas9 gRNS-ére és hogy hova illesztettük a bárkódot. A bárkód szolgált arra, hogy a különböző hosszúságú és mennyiségű RNS-ek kiindulási mennyiségét, valamint a DNS templátjaik szekvenciáját azonosíthassuk. A bárkódok nemcsak abban nyújtottak segítséget, hogy ez alapján beazonosíthassuk melyik RNS szekvencia melyik templátról keletkezett. Azzal, hogy egy szekvencianégyeshez sok teljesen más szekvenciájú 12 bázisos bárkód tartozott, az RNS mennyiségekre ható külső tényezők hatását csökkentettük, mint pl. RNS szerkezetbeli különbségekből adódó eltérő RNS degradációt.



#### 17. Ábra. Randomizált bázisokkal módosított SpCas9 gRNS

Az ábra az SpCas9 gRNS másodlagos szerkezetét ábrázolja a randomizált bázisokkal. X-el azok a nukleotidok vannak jelölve, amelyeknek a helyére illesztettük a bárkódot.

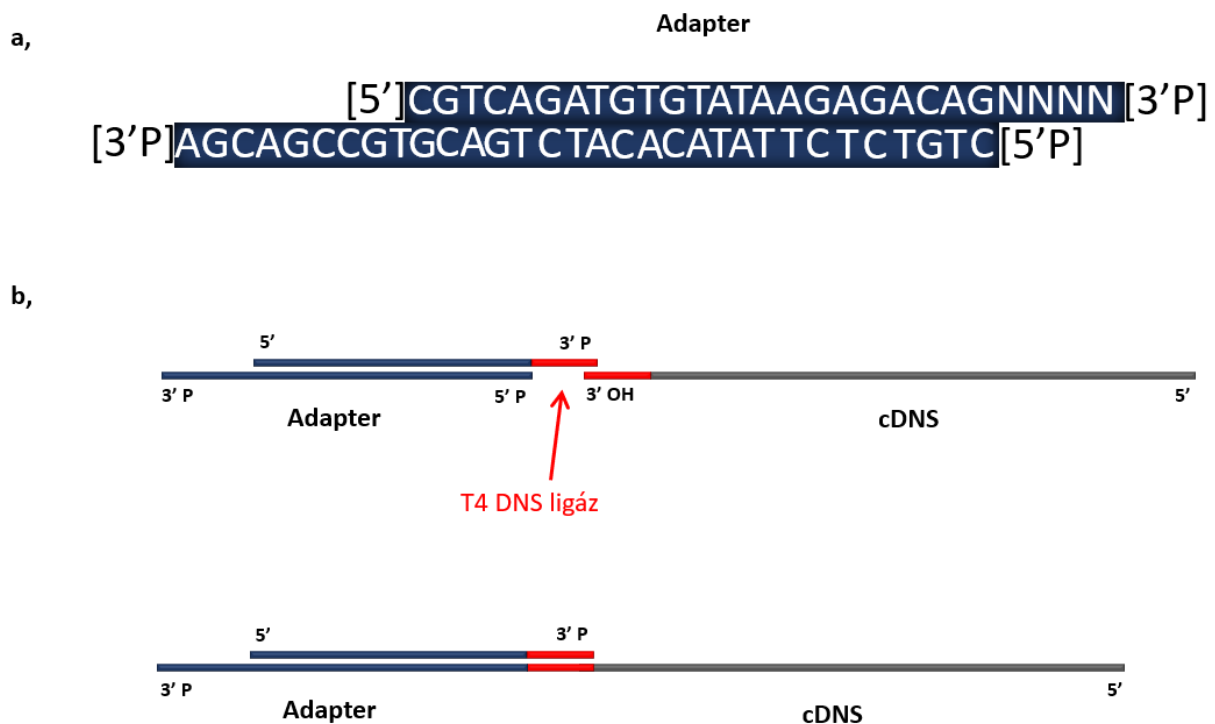
Ábra forrása:

<https://horizondiscovery.com/en/applications/crispr-cas9/synthetic-sgrna-for-crispr-cas9-experiments> alapján

### 5.3.2. cDNS+adapter ligálási módszer

A cDNS-re való adapter ligálásához egyedi módszert alkalmaztunk több okból kifolyólag. A meglévő módszerek olyan RNS-ek szekvenálásához, mint pl. a Cas9 guide RNS, melynek 3' végének szekvenciája ismert, általában felesleges lépéseket tartalmaznak, és emellett túlságosan idő és anyagigényesek. Módszerünk során a tisztított RNS-eket rögtön cDNS-ekké írjuk át, majd a cDNS-t ligáljuk egy ismert szekvenciájú adapterrel, végül a ligátumot PCR-el amplifikáljuk NGS-hez.

Az adapter ligáláshoz a T4 DNS ligázt választottuk, abból az okból kifolyólag, hogy nincs nukleotid preferenciája<sup>76,77</sup>. Bár alapvetően ezt a ligázt duplaszálú DNS-ek tompa, vagy ragadós végeinek összeligálásához használják, egyszálú DNS-ek ligálására is képes egy splint oligó segítségével<sup>76</sup>. Ezek ismeretében azt a módszert választottuk, melynek sematikus ábrázolása a 18. ábrán látható. A cDNS-re ligálandó adapterhez hozzáhibridizálunk egy részben komplementer oligót, aminek van egy 4 randomizált nukleotidot tartalmazó túlnyúló vége, és a cDNS első 4 nukleotidjához annealálódva, a T4 DNS ligáz képes összeligálni az adaptert a cDNS-el. A ligáláshoz használt oligonukleotidokat úgy terveztük meg, hogy ne keletkezhessenek ligálási melléktermékek. Az adapter komplementer oligó 3', és az adapter 3' végén is egy plusz foszfát csoport van, amely gátolja a ligáz működését. Csak az adapter 5' végén van foszfát csoport, amely a T4 ligáz működéséhez szükséges (18. ábra). Mivel a T4 DNS ligáz 5'P véget ligál 3'OH véghez, így csak az adapter 5'P vége tud a cDNS-el ligálódni. Amiatt, hogy a T4 DNS ligáz egyszálú DNS-ek ligálását kisebb hatékonysággal végzi, mint a duplaszálúakét, a módszer eredményes használatához be kellett állítani az optimális kondíciókat. Ezek közé tartozott a ligálás időtartama, hőmérséklete, oligó koncentráció és PEG koncentráció (melynek használata elősegíti a ligálás hatékonyságát). Vizsgálataink alapján optimális körülményeknek a következőket találtuk: 37 °C, 1 h, 20% PEG.

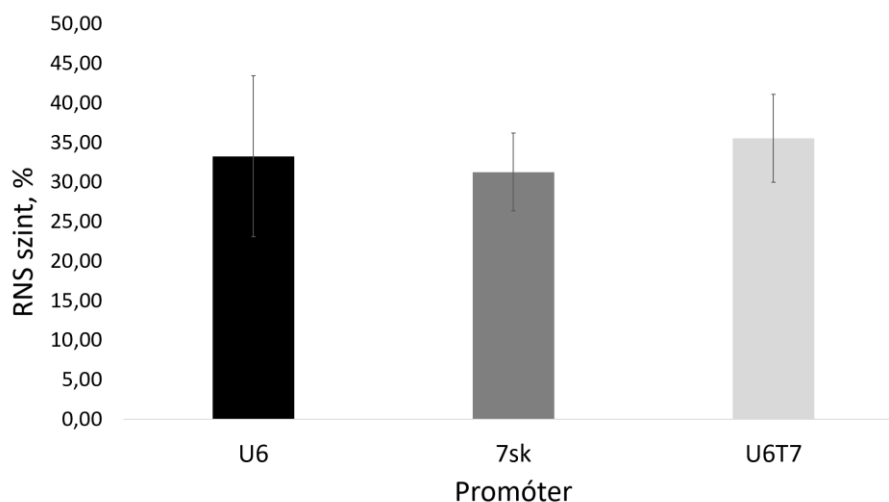


#### 18. Ábra. cDNS+adapter ligálás

Az ábrán a cDNS-re ligált adapter és adapter komplementer oligók szekvenciája (a), valamint a cDNS+adapter ligálás sematikus ábrázolása (b) látható. **a**, A két oligó szekvenciája, végmódosításaikkal. **b**, Pirossal a 4 nt hosszú randomizált nukleotidokat tartalmazó szakasz, szürkével a cDNS, kékkel az adapter és komplementere van ábrázolva.

#### 5.3.3. Emlős promóterek összehasonlítása

Azt kívántuk összehasonlítani, hogy HEK-293 sejtekben a humán U6, humán 7sk, U6T7 hibrid promóterekről mennyi RNS termelődik egymáshoz képest. Ehhez 3 plazmidkönyvtár keveréket csináltunk, ahol keverékenként más spacer szekvenciát tartalmazó gRNS-ek expresszálódnak az egyes promóterekről. A plazmidkeverékeket egyszer csináltuk meg, ezekből 3-3 párhuzamost transzfektáltunk a sejtekbe. Az NGS elemzéséből kapott plazmidkeverékek readszámaira normalizáltuk az RNS-ek readszámaikat, megfelelő promóter-spacer párosonként. Ezután megnéztük, hogy melyik RNS könyvtár mekkora százalékot tett ki az adott keverék összes RNS-éből. Az eredmények alapján, amelyek a 19. ábrán láthatók, nincs szignifikáns eltérés a 3 promóter között, hasonló mennyiségű RNS termelődik mindegyik promóterről HEK-293 sejtekben.

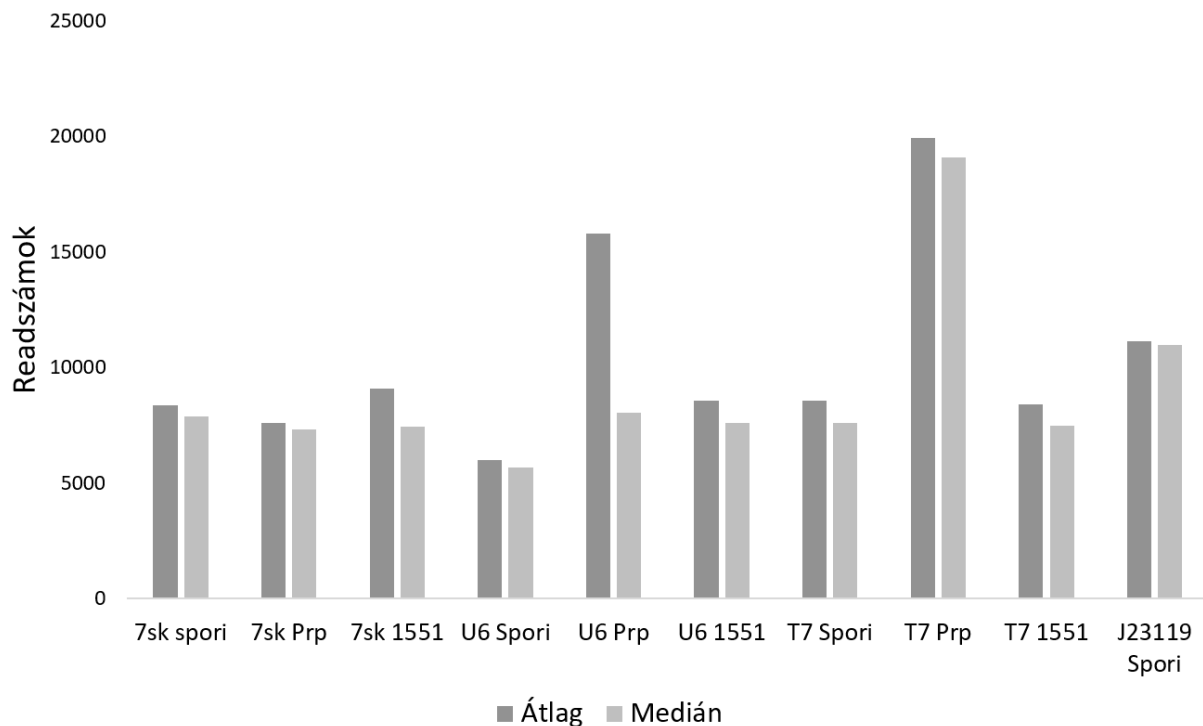


#### 19. Ábra. Emlős promóterek expressziójának összehasonlítása HEK-293 sejtekben

Az ábra megmutatja, hogy az U6, 7sk, U6T7 promóter-gRNS kazettákat tartalmazó plazmidokat kotranszfectálva emlős sejtekbe az egyes promóterekről az összes RNS hány %-a keletkezett. Feketével az U6, sötétszürkével a 7sk, világosszürkével az U6T7 promóterről keletkezett RNS-ek mennyisége látszik. A hibasávok a százalékos értékek átlagának szórását mutatják.

#### 5.3.4. A promótertől downstream elhelyezkedő +1-től +4-ig terjedő pozíciók vizsgálata az RNS expresszióra és a transzkripció iniciációs pozícióra

Annak tanulmányozására, hogy a CRISPR gRNS expresszióra gyakran alkalmazott promóterekről termelődő RNS-ek mennyiségét és hosszát megfigyelhessük, 5 promótert, a humán U6, humán 7sk, az U6T7 hibrid promótert, a T7 és a bakteriális J23119 promótert vizsgáltunk meg. Az NGS adatok feldolgozása alapján az összes plazmid könyvtár promótert követő 4 bázispáros szekvencia variációinak mindegyike (256 variáció) átlagosan minimum 5000 darab számú readdal képviseltette magát (20. ábra).



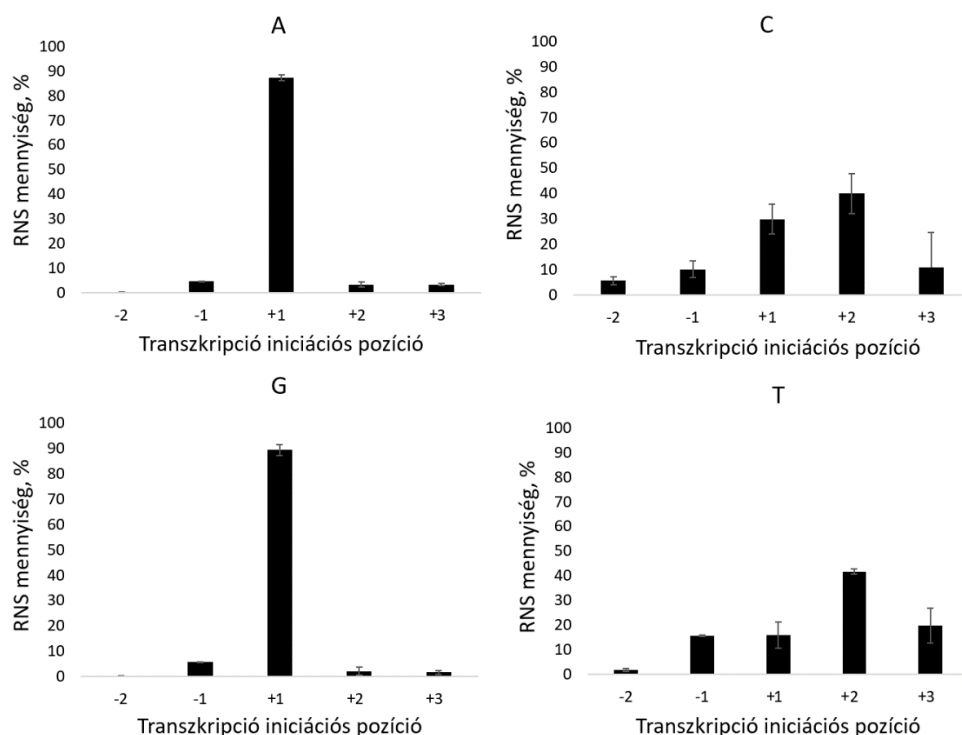
**20. Ábra. A promótert követő 4 bázispáros szekvencia variációk NGS readszámainak átlagai és mediánjai plazmidkönyvtáranként**

Az ábrán a promótert követő 4 bázispáros szekvencia variációk readszámainak átlagai és mediánjai vannak feltüntetve plazmidkönyvtáranként. A sötétszürke oszlopok a readszámok átlagait, míg a világosszürkék a mediánokat mutatják.

A plazmidkönyvtárakból létrehozott cDNS könyvtárak readszámainak mennyisége és a readek hossza, amelyekből az RNS-ek hosszát számoltuk ki, a promótert követő 4 bázisos szekvenciától és a promóterrel függően változott. Az emlős promóterek összehasonlítását kivéve az összes további eredmény az NGS adatokból úgy került kiszámításra, hogy a plazmid könyvtárak bárkódjai alapján azonosítottuk és a templátra normalizáltuk az adott templátokról keletkezett RNS-ek szekvenciáját, 5' végének hosszát és RNS mennyiségét. Ezt követően a különböző bárkódokról keletkezett, eltérő hosszúságú RNS-ek mennyiségét kezdő szekvenciánként és azonos hosszúságonként összeadtuk, cDNS könyvtáranként. Promóterenként 3-féle spacer tartalmazó SpCas9 gRNS 3-3 párhuzamos cDNS könyvtárakat vizsgáltunk, kivéve az U6 promótert, ahol a Spori spacerrel 2 párhuzamos cDNS könyvtár volt, és a J23119 promótert, ahol csak a Spori spacer 3 párhuzamos cDNS könyvtárát vettük figyelembe, abból az okból kifolyólag, hogy a többi esetben nem volt minden kezdő szekvenciáról megfelelő számú read.

### 5.3.4.1. Az U6 promóter

Az NGS eredményeink alapján először megvizsgáltuk az első nukleotid szerepét a transzkripció iniciációs pozíciókra, melynek eredménye a 21. ábrán látható. Ha a +1-es pozícióban A volt, akkor az összes RNS átlag 87%-a keletkezett a +1 pozícióról, +1G-vel ez 89% volt. A purinokhoz képest rendkívül kis arányban indul el pontosan az átírás, ha a +1 pirimidin: +1C-vel 30%, +1T-vel 16% volt a +1-ről átíródott RNS-ek aránya.



**21. Ábra.** Az U6 promótertől downstream elhelyezkedő első nukleotid hatása a különböző transzkripció iniciációs pozíciókra

Az ábrán látható, hogy az U6 promótertől downstream elhelyezkedő első nukleotid miképp befolyásolja, hogy melyik pozícióról milyen arányban indul el az átírás. Egy oszlop megmutatja, hogy az összes vizsgált pozícióról termelődött RNS-ek összegének átlagban hány %-a íródott át adott pozícióról. A hibasávok a százalékos értékek átlagának szórását mutatják. N=8, amely a 8 cDNS könyvtárnak felel meg.

Ezután megvizsgáltuk a további pozíciókban lévő nukleotidok szerepét is a különböző transzkripció kezdő pozíciókra (22. ábra). Azt találtuk, ha az első pozícióban purin van, a további pozíciók már nem befolyásolják olyan nagymértékben a kezdő pozíciót, mint a kezdő pirimidineknél, hanem az RNS-ek nagyrészt a +1-ről íródnak át. Ha a +1-ben pirimidin van, és a +2-ben purin, akkor nagyobb arányban indul el a +2-ről a transzkripció, mint a +1-ről, ilyenkor kimarad a +1, és az átírás az első purint tartalmazó nukleotidról, a +2-es pozícióról indul el. Ha a +1 és a +2 is pirimidin, de a +3 purin, szintén megfigyelhető, hogy megnő azoknak az RNS-eknek az aránya, ahol kimarad a +1 és a +2 is, és az átírás az első purint tartalmazó nukleotiddal, a +3-as pozícióról indul el.

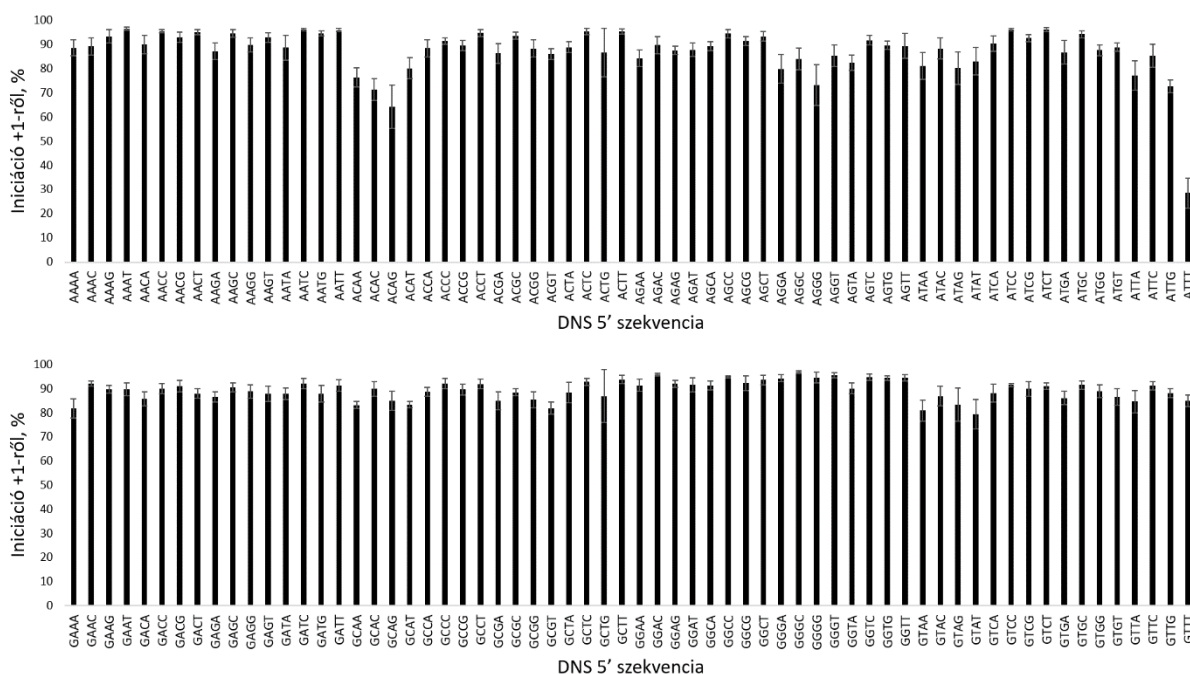




## 22. Ábra. Az U6 promótertől downstream elhelyezkedő szekvenciák hatása a különböző transzkripció iniciációs pozíciókra

Az ábrán látható, hogy az U6 promótertől downstream elhelyezkedő szekvenciák miképp befolyásolják, hogy melyik pozícióról milyen arányban indul el az átírás. Egy oszlop egy színe megmutatja, hogy az összes vizsgált pozícióról termelődött RNS-ek összegének átlagban hány %-a íródott át adott pozícióról, és egy oszlop egy darab kezdő szekvenciát jelöl. A közepesen világoskék szín a -3, a narancssárga szín a -2, a szürke szín a -1, a citromsárga szín a +1, a világoskék szín a +2, a zöld szín a +3 és a sötétkék szín a +4 pozíciókat jelöli az U6 promótertől downstream. N=8, amely a 8 cDNS könyvtárnak felel meg.

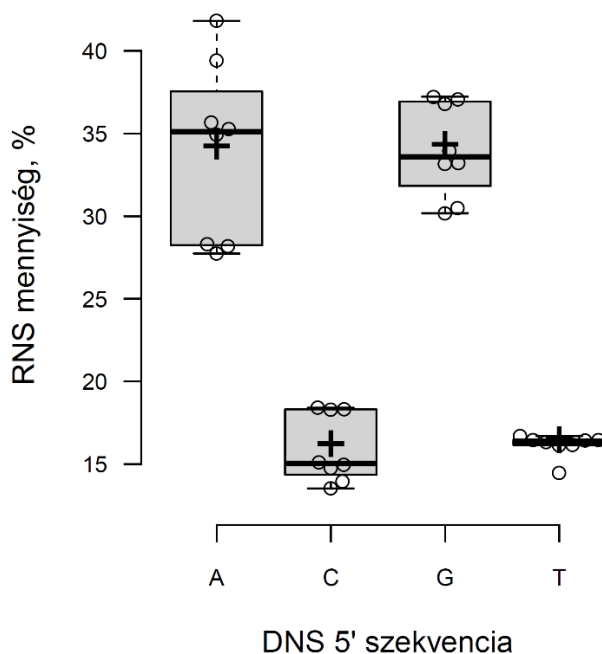
A+1-es pozícióban A-val, vagy G-vel szinte csak jó hosszúságú RNS-ek keletkeztek, legnagyobb arányban a +1-ben kezdődik az átírás (23. ábra). Van néhány szekvencia +1A-val, amelyiknél a +1-es pozícióról induló transzkripció aránya kisebb, de +1A és +1G között nincs jelentős eltérés abban, hogy milyen arányban indul el a transzkripció iniciációja a +1-ről. Eredményeink azt mutatják, hogy ha az U6 promóter után az első nukleotid purin, az átírás nagyrészt a +1-ről indul el, és a kezdő pozíciót a további pozíciók csak kis mértékben befolyásolják, míg ha a +1 pirimidin, az átírás nemcsak +1-ről történik. Azt, hogy pirimidin esetében melyik pozícióról milyen arányban történik a transzkripció iniciációja, nagymértékben befolyásolják a promótertől downstream elhelyezkedő további nukleotidok is.



## 23. Ábra. Az U6 promótertől downstream lévő szekvenciák hatása a +1-ről induló transzkripcióra, ha a +1 A, vagy G

Az ábrán az U6 promóter után elhelyezkedő különböző szekvenciák szerepe látható a +1-es pozícióról történő átírásra. Az oszlopok megmutatják, hogy átlagban hány %-ban indul el +1-ről az átírás adott templát DNS szekvencia esetén. A hibaszávok a százalékos értékek átlagainak szórását mutatják. N=8, amely a 8 cDNS könyvtárnak felel meg.

A következő lépésekben azt néztük meg, hogy az U6 promótertől downstream elhelyezkedő nukleotidok hogyan hatnak a transzkriptumok mennyiségére. A +1-es pozíció szerepét vizsgálva az átíródott RNS-ek mennyiségére azt találtuk, hogy ha a +1-ben purin van, több RNS termelődik, mintha pirimidin (24. ábra). A purinokról és a pirimidinekről keletkezett transzkriptumok mennyisége egymástól szignifikánsan eltér, a 2 purint egymással, illetve a 2 pirimidint egymással összehasonlítva azonban nincs szignifikáns különbség.

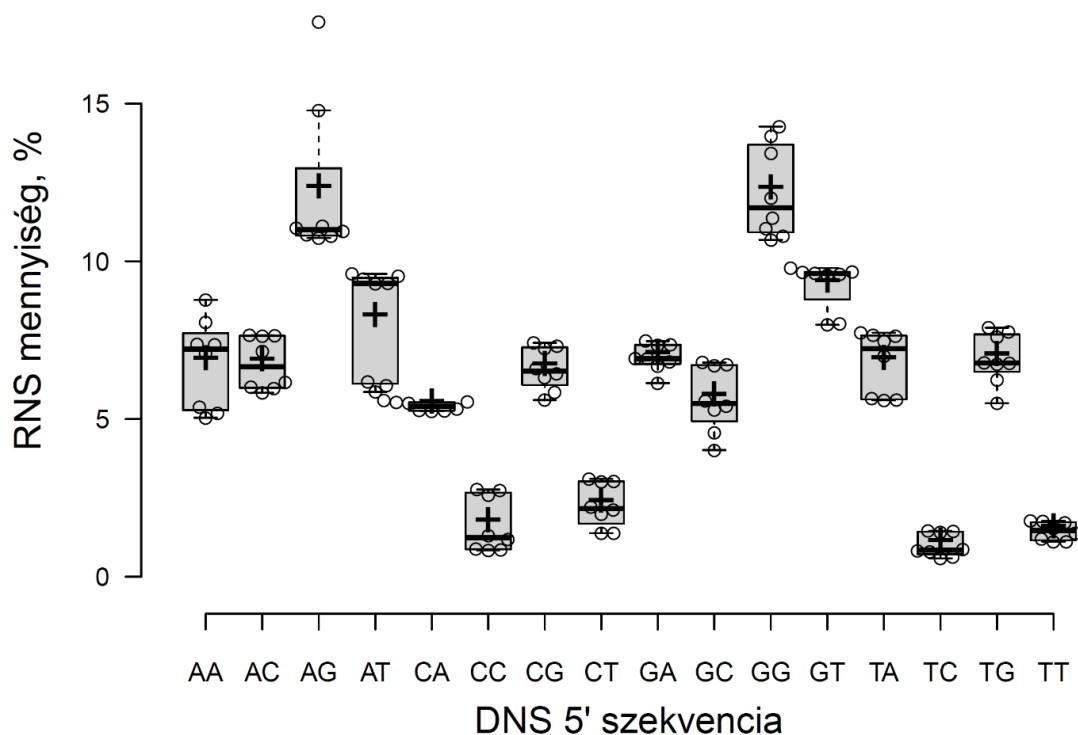


**24. Ábra. A +1-es pozícióban lévő nukleotid hatása az RNS expresszióra**

Az ábrán a +1-es pozícióban lévő nukleotid hatása látható az U6 promóterről termelődött RNS-ek mennyiségére. Az ábra az összes transzkripció iniciációs pozíció transzkriptumainak mennyiségét mutatja. 1 adatpont megadja, hogy adott nukleotiddal az összes RNS hány %-a keletkezett az adott cDNS könyvtáron belül. A boxokon belüli vastag vízszintes vonalak a mediánokat mutatják; a boxok határolják az R programmal meghatározott 25. és 75. percentilist; a szimpla vízszintes vonalak az interkvartilis tartomány 1,5-szereséig terjednek ki a 25. és 75. percentilistől; a keresztek a minták átlagát mutatják; az adatpontok körökként vannak ábrázolva. N=8, amely a 8 cDNS könyvtárnak felel meg. A p értékek a Függelékben a 109. oldalon találhatóak meg.

A promótert követő első 2 pozíciót vizsgálva azt kaptuk, hogy a legtöbb RNS AG, GG kezdettel keletkezik (25. ábra). AG és GG szignifikánsan a legjobbak a 16 variáció közül és egymástól nem különböznek. CC, CT, TC, TT pedig szignifikánsan mindennél kevesebb RNS-t eredményeznek (25. ábra). Összehasonlítottuk egymással a +1 pozícióban A-t tartalmazó templátokról termelődött RNS-ek mennyiségét. Az első két pozíciót vizsgálva az látszik, hogy szignifikáns különbség csak az AG és az összes többi között van, AA, AC, AT között nincs, és AG kezdettel termelődött a legtöbb RNS. Az első két pozíciót vizsgálva a GA-GC párost leszámítva minden variáció szignifikáns különbségeket okozott +1G-vel az RNS mennyiségekben, és a 2. pozícióban a G után is a G bizonyult a legjobbnak (25. ábra). A 25. ábrán látható, hogy az első két pozíció dinukleotidjai különösen meghatározóak az RNS

expresszióra nézve: AG és a GG szignifikánsan nagyobb, a legkisebb expressziójú +1-ben purint tartalmazó GC szekvenciák kétszeresét elérő expressziót biztosít. Eredményünk alapján látható, hogy az U6 promóternél a 2. nukleotid is rendkívül nagymértékben befolyásolja a keletkező RNS-ek mennyiségét, transzkriptálódhat olyan mennyiségű RNS a +1-es pozícióban pirimidinről is, mint purinról, ha megfelelő nukleotid követi, valamint a +1A-val, vagy +1G-vel keletkezett RNS-ek mennyiségében is lehet jelentős eltérés, ha a második pozícióban különböző nukleotidok szerepelnek.

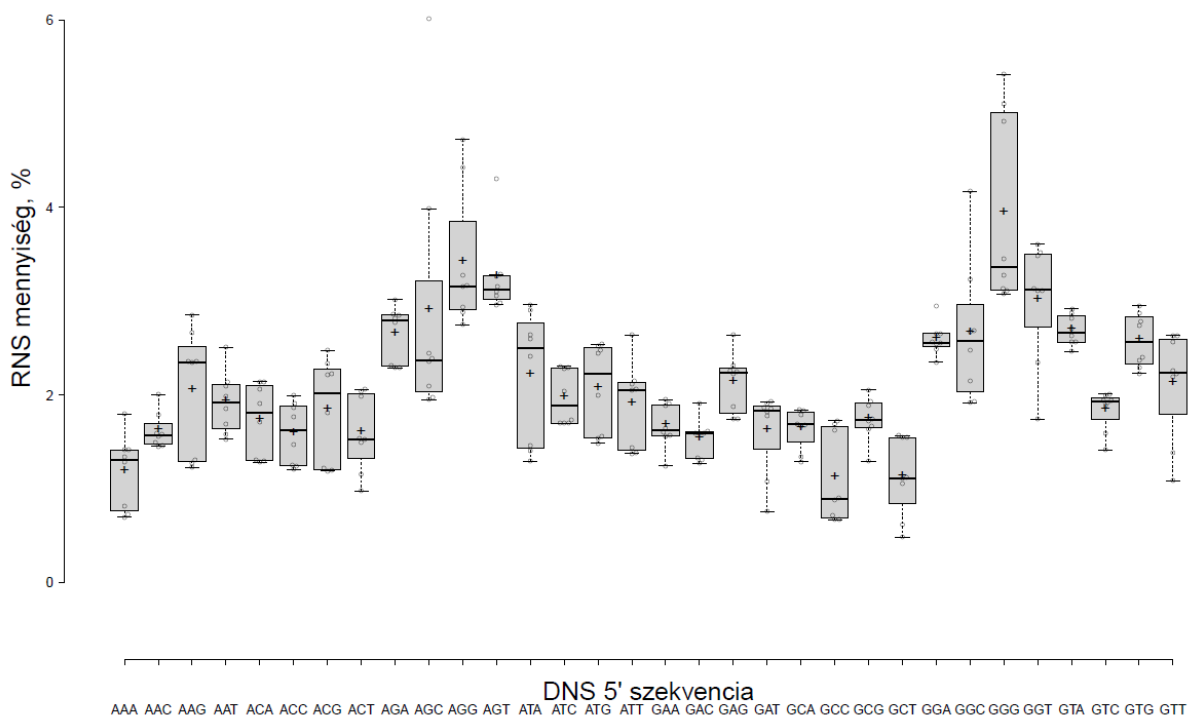


**25. Ábra. A promótertől downstream elhelyezkedő első két nukleotid hatása az RNS expresszióra**

Az ábrán a promótert követő első két nukleotid hatása látható az U6 promóterről termelődött RNS-ek mennyiségére. Az ábra az összes transzkripció iniciációs pozíció transzkriptumainak mennyiségeit mutatja. 1 adatpont megadja az összes szekvenciáról termelődött RNS-ek összegéhez képest keletkezett RNS-ek százalékos arányát az adott szekvenciáról, adott cDNS könyvtárban belül. A boxokon belüli vastag vízszintes vonalak a mediánokat mutatják; a boxok határolják az R programmal meghatározott 25. és 75. percentilist; a szimpla vízszintes vonalak az interkvartilis tartomány 1,5-szereséig terjednek ki a 25. és 75. percentilistől; a keresztek a minták átlagát mutatják; az adatpontok körökként vannak ábrázolva. N=8, amely a 8 cDNS könyvtárnak felel meg. A p értékek a Függelékben a 109. oldalon találhatók meg.

Mivel az eddig bemutatott eredmények alapján látható, hogy ha a +1-es pozícióban pirimidin van, pirimidinekkal a transzkripció iniciációs pozíciók sokkal heterogénebbek, ezért a további pozíciók szerepét az RNS mennyiségekre már csak +1A-val és +1 G-vel vizsgáltuk a továbbiakban. Az első 3 pozíció szekvenciáját vizsgálva azt kaptuk, hogy a legtöbb szignifikáns eltérés az AG kezdetűek és a többiek között van. AGT, AGG mindnél szignifikánsan jobb volt, kivéve a többi AG kezdetű variációnál (26. ábra).

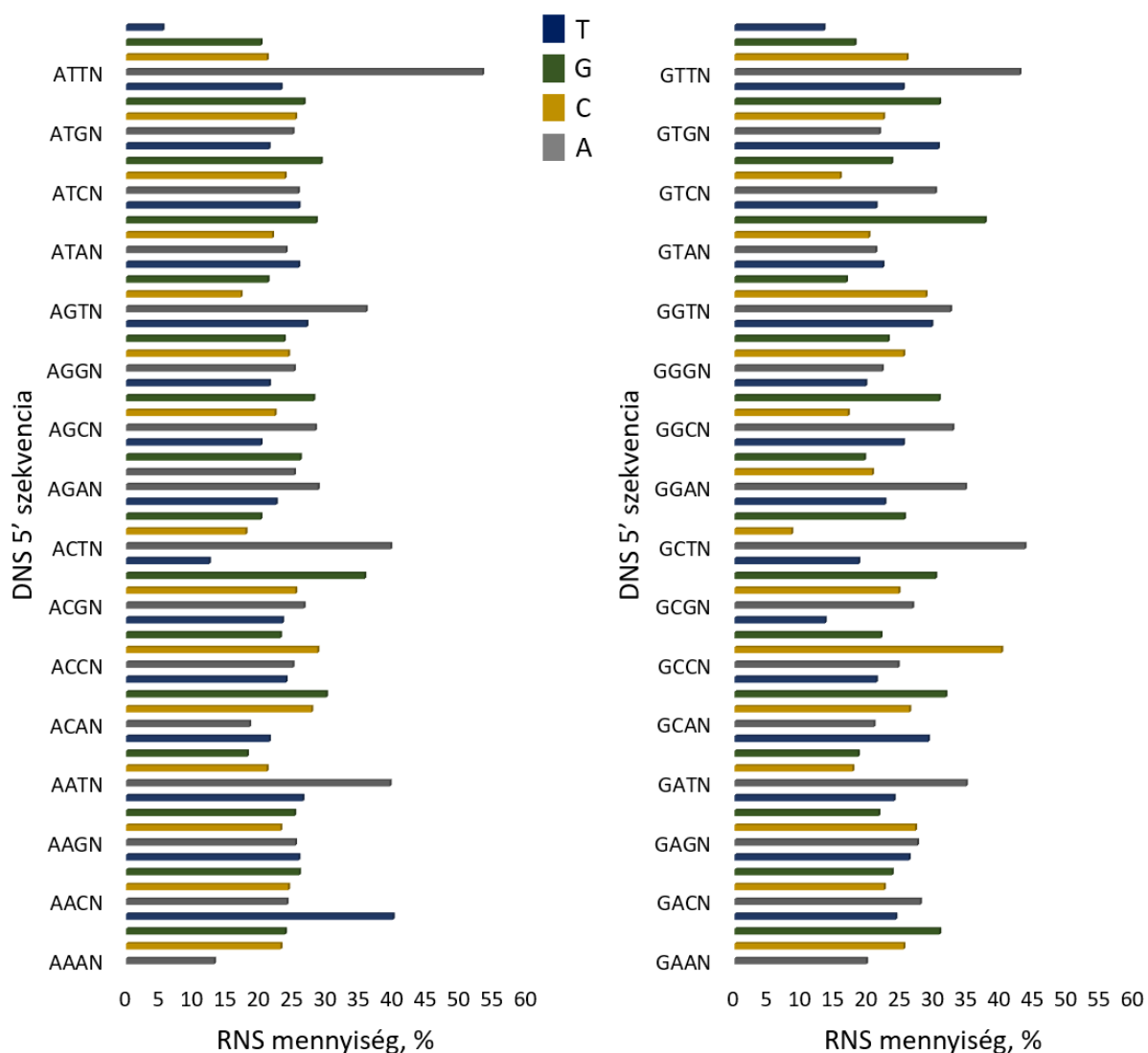
Az első 3 pozíció szekvenciáját vizsgálva +1 G-vel nagyobb különbségeket találtunk. A legtöbb különbség a GG kezdetű dinukleotidok és a többi variációjú dinukleotiddal kezdődő szekvenciák között van, és a GGG-vel kezdődő szekvenciák majdnem minden más szekvencia variációnál szignifikánsan több RNS-t eredményeznek. A GT kezdetű szekvenciákról keletkezett RNS mennyiségekben a +3-as pozíció nukleotidjai is szignifikáns különbséget okoztak (26. ábra). Eredményeink alapján, ha a +1-es és a +2-es pozíciókban lévő dinukleotid megegyezik, viszont az azt követő pozíció nukleotidja más, az is különbségeket okozhat az RNS mennyiségekben.



**26. Ábra. A promótertől downstream elhelyezkedő első három nukleotid hatása az RNS expresszióra, ha a +1-es pozícióban A vagy G van**

Az ábrán a promótert követő első három nukleotid hatása látható az U6 promóterről termelődött RNS-ek mennyiségére, ha az első pozícióban A, vagy G van. Az ábra az összes transzkripció iniciációs pozíció transzkriptumainak mennyiségeit mutatja. 1 adatpont megadja az adott szekvenciáról keletkezett RNS-ek százalékos arányát az összes szekvenciáról termelődött RNS-ek összegéhez képest az adott cDNS könyvtárban. A boxokon belüli vastag vízszintes vonalak a mediánokat mutatják; a boxok határolják az R programmal meghatározott 25. és 75. percentilist; a szimpla vízszintes vonalak az interkvartilis tartomány 1,5-szereséig terjednek ki a 25. és 75. percentilistől; a keresztek a minták átlagát mutatják; az adatpontok körökként vannak ábrázolva. N=8, amely a 8 cDNS könyvtárnak felel meg. A p értékek a Függelékben a 110-111. oldalon találhatók meg.

A +4-es pozíció hatását is megvizsgáltuk az RNS expresszióra, és azt találtuk, hogy akár néhányszoros RNS mennyiségbeli különbségeket is képes okozni a 4. pozícióban elhelyezkedő bázis, ha az első 3 megegyezik, tehát még a +4-es pozícióban lévő nukleotid is befolyásolja a transzkripciót. A 27. ábrán látható, hogy a +4-es pozíció nem egyforma módon befolyásolja a transzkriptum mennyiségét, hanem a tőle upstream lévő szekvencia függvényében változóan.

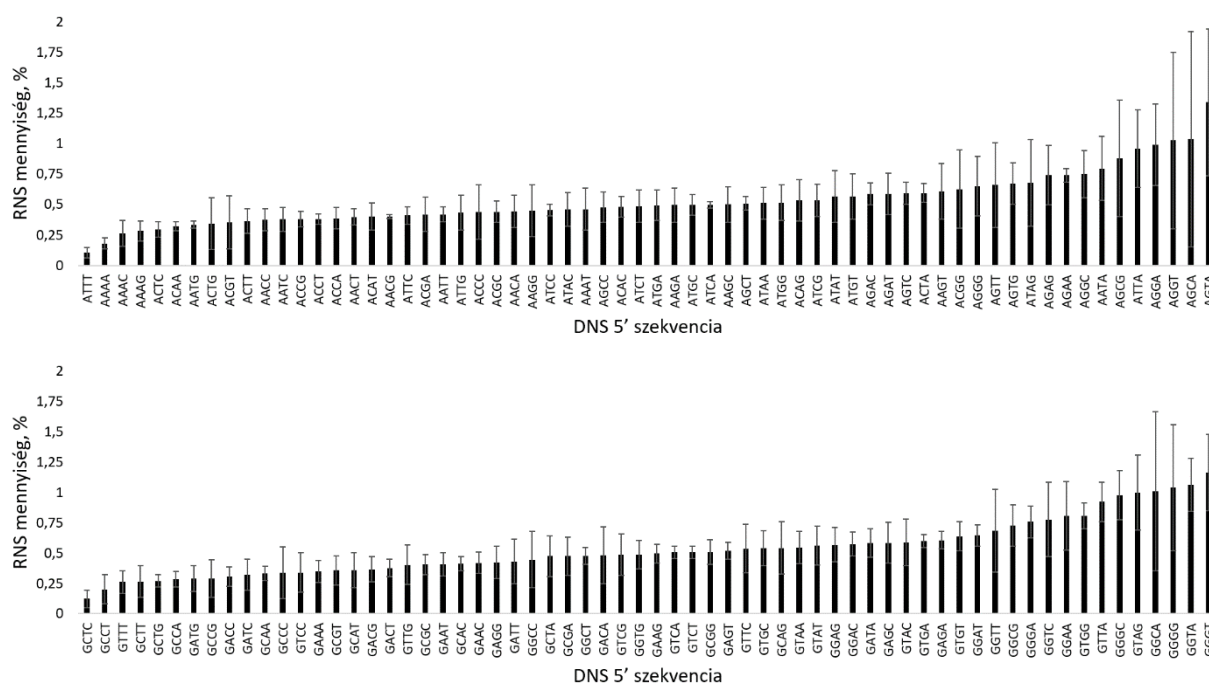


**27. Ábra. A +4-es pozícióban lévő nukleotidok hatása az RNS expresszióra, ha a +1 A, vagy G**

Az ábrán a +4-es pozícióban elhelyezkedő különböző nukleotidok hatása látható az U6 promóterről termelődött RNS-ek mennyiségére, ha a +1-es pozícióban A vagy G van. Az ábra az összes transzkripció iniciációs pozíció transzkriptumainak mennyiségeit mutatja. 1 oszlop megadja, hogy a promótert követő 4 bázis adott szekvenciájáról, 8 db cDNS könyvtár alapján átlagosan hány % RNS keletkezett az összes, első 3 pozíció szekvenciájában megegyező, de 4. pozícióban más nukleotidot tartalmazó szekvenciáról termelődött RNS-ek összegéhez képest.

A +1-es pozíciótól a +4-esig terjedő szekvencia függvényében a fenti ábrákon látható módon változó mennyiségű RNS expresszálódik az U6 promóterről. A legkisebb és a legnagyobb expressziót biztosító szekvenciákról termelődött RNS mennyiségek átlagai között közel 13x-os különbség van +1A-val (ATTT és AGTA) és több, mint 9x-es +1G-vel (GCTC és GGGT) (28. ábra).

Eredményeink alapján az U6 promóternél nemcsak a +1-es pozíció, hanem a továbbiak is befolyásolják a keletkező RNS-ek mennyiségét.

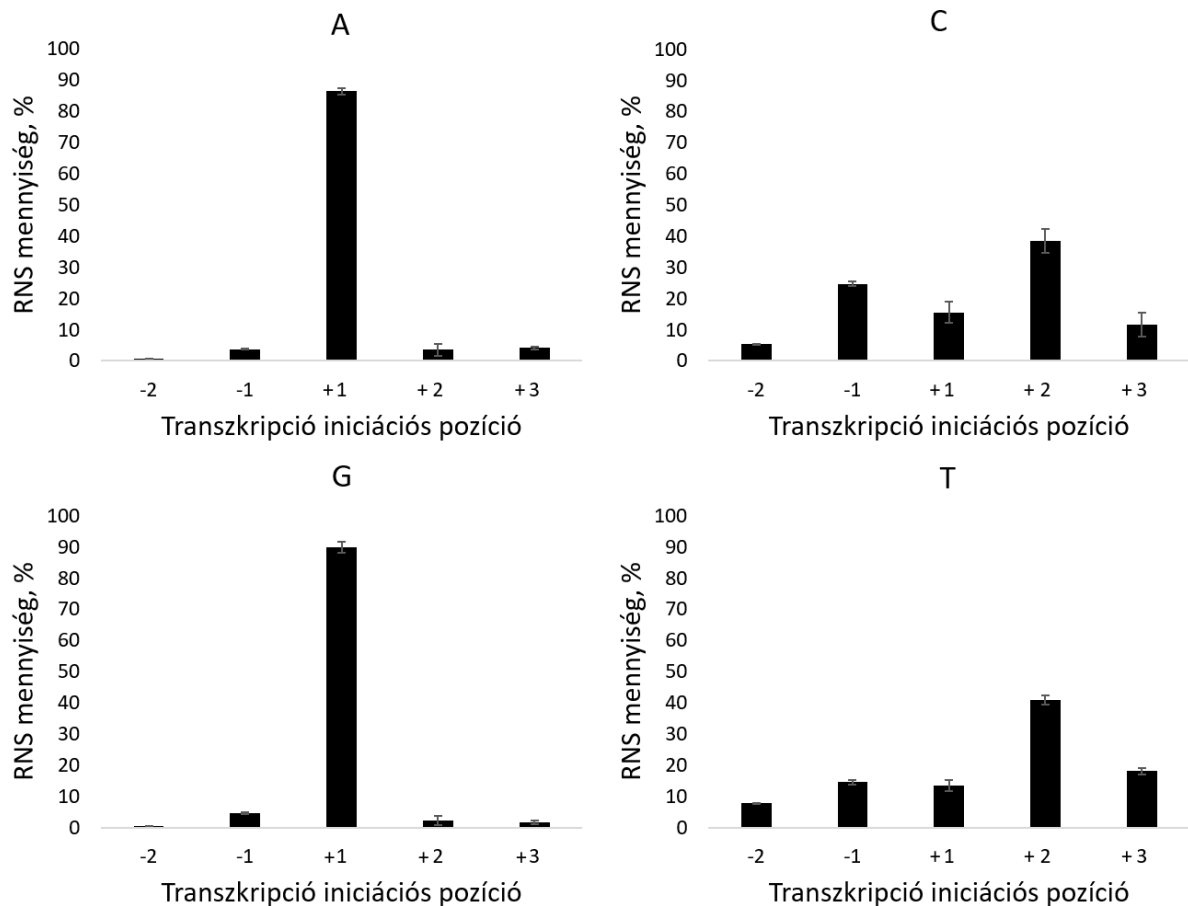


**28. Ábra.** Az U6 promótert downstream követő négy bázisos szekvencia hatása az RNS expresszióra, ha a +1 A vagy G

Az ábrán a +1-től a +4-es pozícióig terjedő szekvencia hatása látható az U6 promóterről termelődött RNS-ek mennyiségére, ha az első pozícióban A vagy G van. Az ábra az összes transzkripció iniciációs pozíció transzkriptumainak mennyiségeit mutatja. 1 oszlop megadja, hogy az összes lehetséges szekvenciáról (256 variáció) termelődött RNS-ek összegéhez képest adott szekvenciáról az RNS-ek hány %-a keletkezett átlagosan. A hibaszávok a százalékos értékek átlagának szórását mutatják. N=8, amely a 8 cDNS könyvtárnak felel meg.

#### 5.3.4.2. A 7sk promóter

Az NGS eredményeink alapján először megvizsgáltuk az első nukleotid szerepét a transzkripció iniciációs pozíciókra (29. ábra). Ha a +1-ben A volt, akkor az összes RNS átlag 86%-a keletkezett a +1 pozícióról, +1G-vel ez 89% volt. Rendkívül kis arányban indul el pontosan az átírás, ha a +1-ben pirimidin volt, a purinokhoz képest. +1C-vel 16%, +1T-vel 13% volt a +1-ről átíródott RNS-ek aránya.

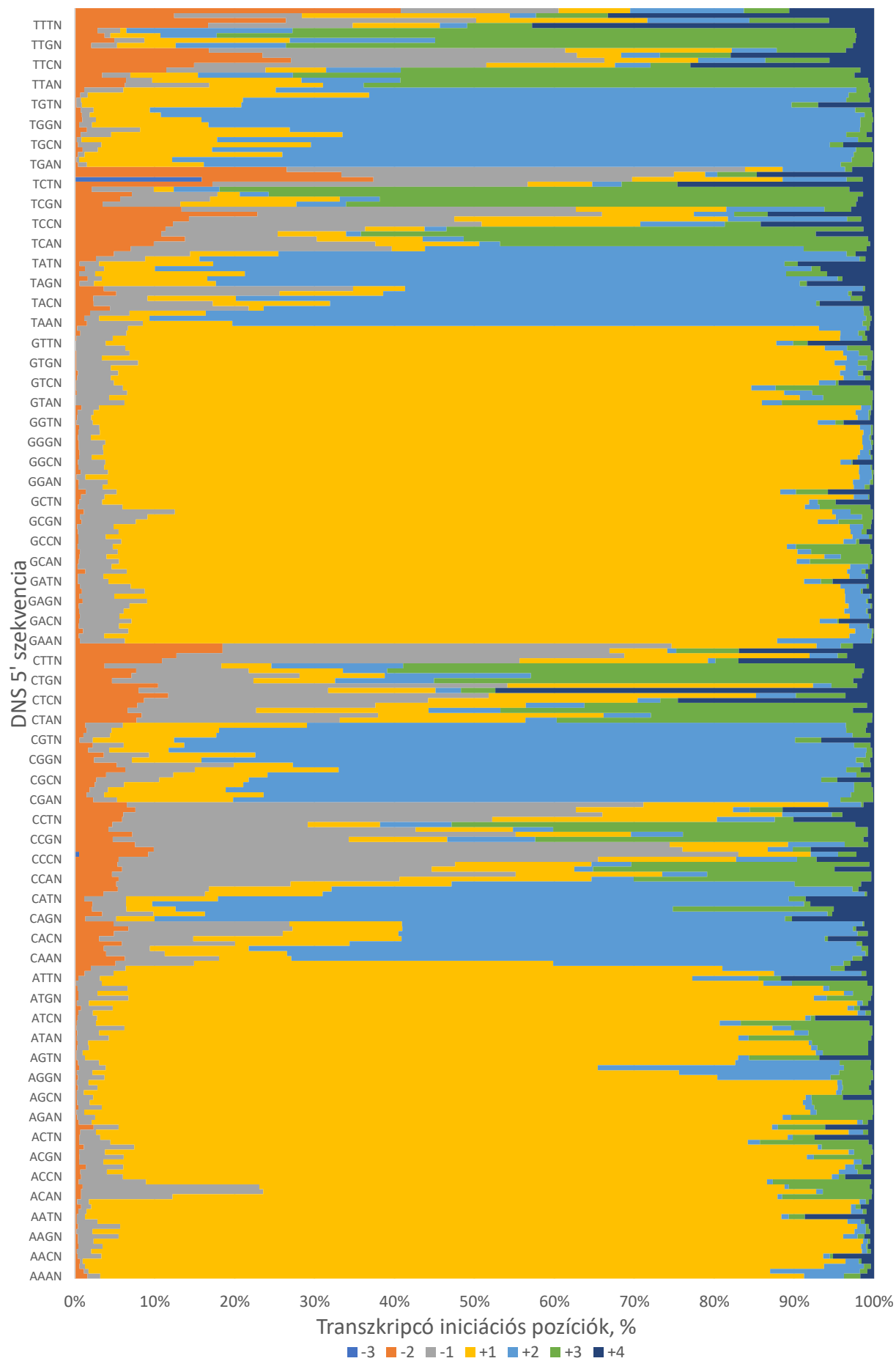


**29. Ábra. A 7sk promótertől downstream elhelyezkedő első nukleotid hatása a különböző transzkripció iniciációs pozíciókra**

Az ábrán látható, hogy a 7sk promótertől downstream elhelyezkedő első nukleotid miképp befolyásolja, hogy melyik pozícióról milyen arányban indul el az átírás. Egy oszlop megmutatja, hogy az összes vizsgált pozícióról termelődött RNS-ek összegének átlagban hány %-a íródott át adott pozícióról. A hibásávok a százalékos értékek átlagának szórását mutatják. N=9, amely a 9 cDNS könyvtárnak felel meg.

Ezután megvizsgáltuk a további pozíciókban lévő nukleotidok szerepét is a különböző transzkripció kezdő pozíciókra (30. ábra). Azt találtuk, ha a +1-es pozícióban purin van, a további pozíciók már nem befolyásolják olyan nagymértékben a transzkripció kezdő pozíciót, az RNS-ek legnagyobbbrészt a +1-ről íródnak át. Ha a +1-es pozíció első bázisa pirimidin, és a +2 purin, akkor nagyobb arányban indul el a +2-ről a transzkripció, mint a +1-ről, ilyenkor kimarad a +1-es nukleotid, és az átírás a következőről indul el. Ha az első 2 bázis pirimidin, de a harmadik purin, szintén megfigyelhető, hogy megnő azoknak az RNS-eknek az aránya, ahol kimaradnak, nem íródnak át a kezdő pirimidint tartalmazó nukleotidok, hanem az átírás az első purint tartalmazó nukleotidról indul el. De ebben az esetben azoknak az RNS-eknek is megnő az aránya, amelyek 1 nukleotiddal hosszabbak és a -1-ről indulnak el (főleg +1C-vel).

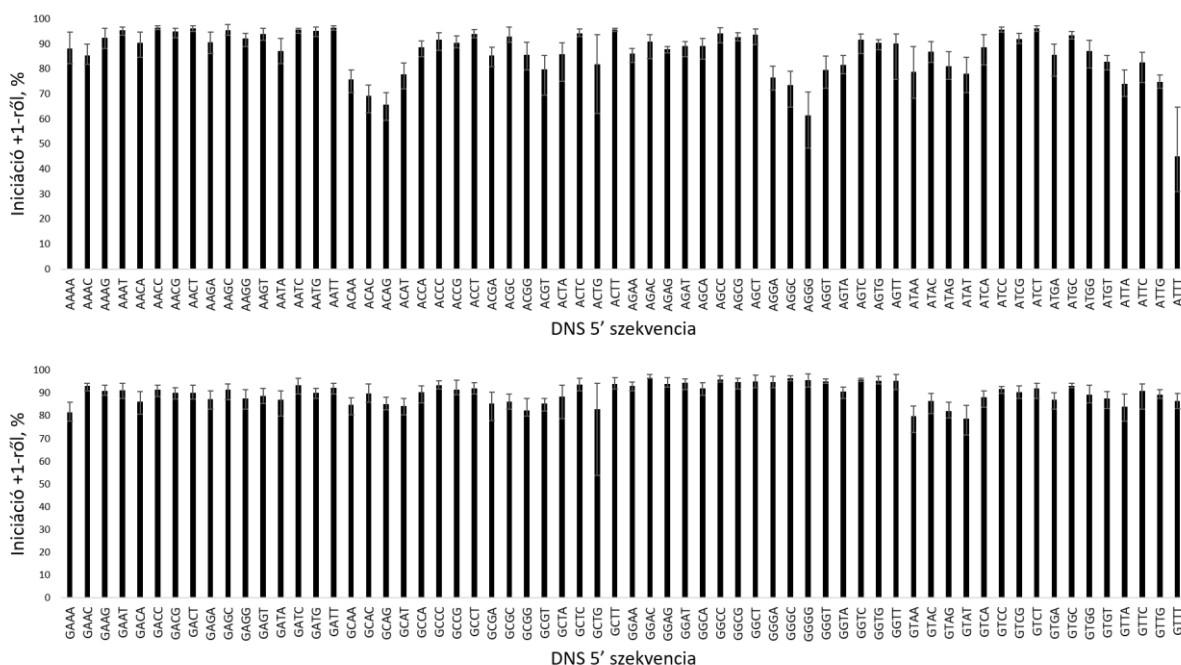




### 30. Ábra. A 7sk promótertől downstream elhelyezkedő szekvenciák hatása a különböző transzkripció iniciációs pozíciókra

Az ábrán látható, hogy a 7sk promótertől downstream elhelyezkedő szekvenciák miképp befolyásolják, hogy melyik pozícióról milyen arányban indul el az átírás. Egy oszlop egy színe megmutatja, hogy az összes vizsgált pozícióról termelődött RNS-ek összegének átlagban hány %-a íródott át adott pozícióról, és egy oszlop egy darab kezdő szekvenciát jelöl. A közepesen világoskék szín a -3, a narancssárga szín a -2, a szürke szín a -1, a citromsárga szín a +1, a világoskék szín a +2, a zöld szín a +3 és a sötétkék szín a +4 pozíciókat jelöli a 7sk promótertől downstream. N=9, amely a 9 cDNS könyvtárnak felel meg.

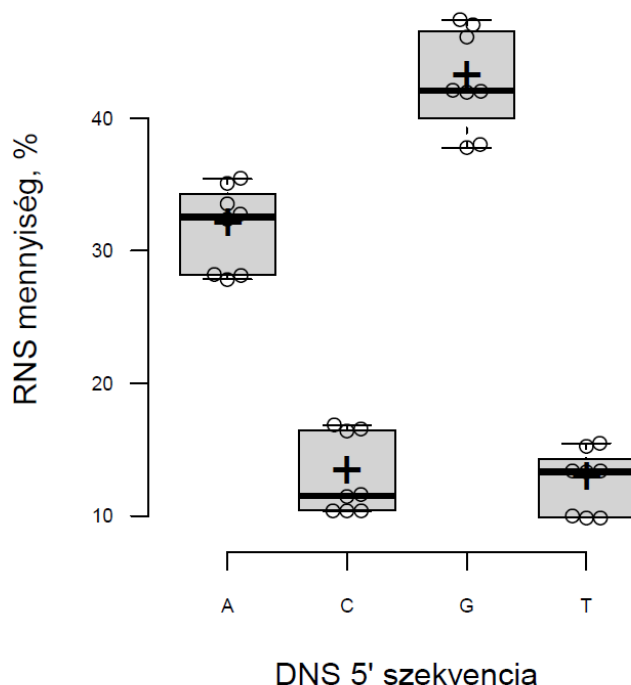
A +1-es pozícióban A-val, vagy G-vel szinte csak jó hosszúságú RNS-ek keletkeztek, legnagyobb arányban a +1-ben kezdődik az átírás (31. ábra). Vannak olyan szekvenciák +1A-val, amelyiknél a +1-es pozícióról induló transzkripció aránya kevesebb, de +1A és +1G között nincs jelentős eltérés abban, hogy milyen arányban lesz a transzkripció iniciációs pozíció a +1. Eredményeink azt mutatják, hogyha a 7sk promóter után az első pozícióban purin van, az átírás majdnem 100%-ban a +1-ről indul el, és a kezdő pozíciót a további pozíciók nagyon kis mértékben befolyásolják. Ezzel szemben, ha a +1-ben pirimidin van, az átírás nemcsak +1-ről történik, és azt, hogy pirimidin esetében melyik pozícióról milyen arányban történik a transzkripció iniciációja, nagymértékben befolyásolják a promótertől downstream elhelyezkedő további nukleotidok is.



### 31. Ábra. A 7sk promótertől downstream lévő szekvenciák hatása a +1-ről induló transzkripcióra, ha a +1 A vagy G

Az ábrán a 7sk promóter után elhelyezkedő különböző szekvenciák szerepe látható a +1 pozícióról történő átírással. Az oszlopok megmutatják, hogy átlagban hány %-ban indult el +1-ről az átírás adott templát DNS szekvencia esetén. A hibaszávok a százalékos értékek átlagainak szórását mutatják. N=9, amely a 9 cDNS könyvtárnak felel meg.

A következő lépésekben azt néztük meg, hogy a 7sk promótertől downstream elhelyezkedő nukleotidok hogyan hatnak a transzkriptumok mennyiségére. Az első pozíció szerepét vizsgálva az átíródott RNS-ek mennyiségére azt találtuk, ha +1-ben purin van, szignifikánsan több RNS termelődik, mintha pirimidin. A purinokról és a pirimidinekről keletkezett transzkriptumok mennyisége egymástól szignifikánsan eltér, a G szignifikánsan jobb, mint az A, kicsit több, mint 30%-al. A 2 pirimidint egymással összehasonlítva viszont nincs szignifikáns különbség (32. ábra).

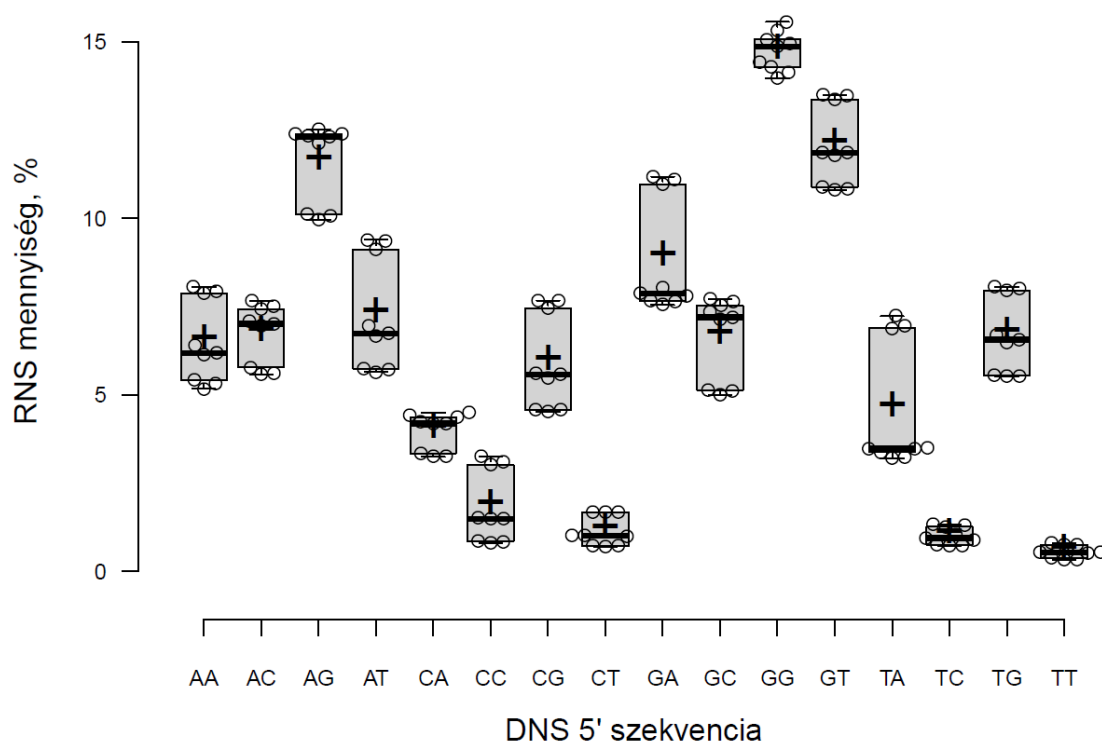


### 32. Ábra. A +1-es pozícióban lévő nukleotid hatása az RNS expresszióra

Az ábrán a +1-es pozícióban lévő nukleotid hatása látható a 7sk promóterről termelődött RNS-ek mennyiségére. Az ábra az összes transzkripció iniciációs pozíció transzkriptumainak mennyiségeit mutatja. 1 adatpont megadja, hogy adott nukleotiddal az összes RNS hány %-a keletkezett az adott cDNS könyvtáron belül. A boxokon belüli vastag vízszintes vonalak a mediánokat mutatják; a boxok határolják az R programmal meghatározott 25. és 75. percentilist; a szimpla vízszintes vonalak az interkvartilis tartomány 1,5-szereséig terjednek ki a 25. és 75. percentilistől; a keresztek a minták átlagát mutatják; az adatpontok körökként vannak ábrázolva. N=9, amely a 9 cDNS könyvtárnak felel meg. A p értékek a Függelékben a 112. oldalon találhatóak meg.

A 7sk promótertől downstream elhelyezkedő első 2 nukleotidot vizsgálva azt kaptuk, hogy a legtöbb RNS GG, AG, GT kezdettel keletkezik, és GG szignifikánsan mindegyiknél jobb, AG és GT között pedig nincs szignifikáns különbség (33. ábra). AG szignifikánsan több RNS-t eredményez, mint a többi olyan variáció, ahol a +1A, illetve, GG is szignifikánsan jobb, minden egyéb +1 G variációnál, sőt a többi 15 variációnál is. +1 G-vel az összes többi verzió szignifikánsan különbözik egymástól, azonban +1A-val csak az AG szignifikánsan jobb a többinél, de azok egymástól nem különböznek.

GA és GT szignifikánsan több RNS-t eredményez, mint az összes +1A, kivéve AG-t. GC-nél csak az AG szignifikánsan jobb, a többi +1A variációtól nem különbözik szignifikánsan. AG szignifikánsan jobb, mint az összes +1C és +1T, de AG kivételével az összes +1A verzió között és a CG, TG közt nincs szignifikáns különbség az RNS mennyiségekben. A legrosszabb G kezdetű (GC) nem különbözik szignifikánsan a legjobb C kezdetűtől (CG) és a legjobb T kezdetűtől (TG), de az összes többi G kezdetű jobb az összes többi C és T kezdetű verziónál. CC, CT, TC, TT szignifikánsan mindennél kevesebb RNS-t eredményeznek, csak egymásnál nem (33. ábra). A +1-ben purint tartalmazó három legnagyobb RNS mennyiséget okozó kezdő dinukleotid mindegyikéről minimum kétszer magasabb RNS szint érhető el átlagban, mint a legalacsonyabb RNS szintet biztosító purin kezdetű dinukleotidról. Eredményünk alapján a 7sk promóternél a +2-es pozíció is rendkívül nagymértékben befolyásolja a keletkező RNS-ek mennyiségét, transzkriptálódhat olyan mennyiségű RNS +1pirimidinről is, mint purinról, ha megfelelő nukleotid követi, valamint a +1A-val, vagy +1G-vel keletkezett RNS-ek mennyiségében is lehet jelentős eltérés, ha a +2-es pozícióban különböző nukleotidok vannak.

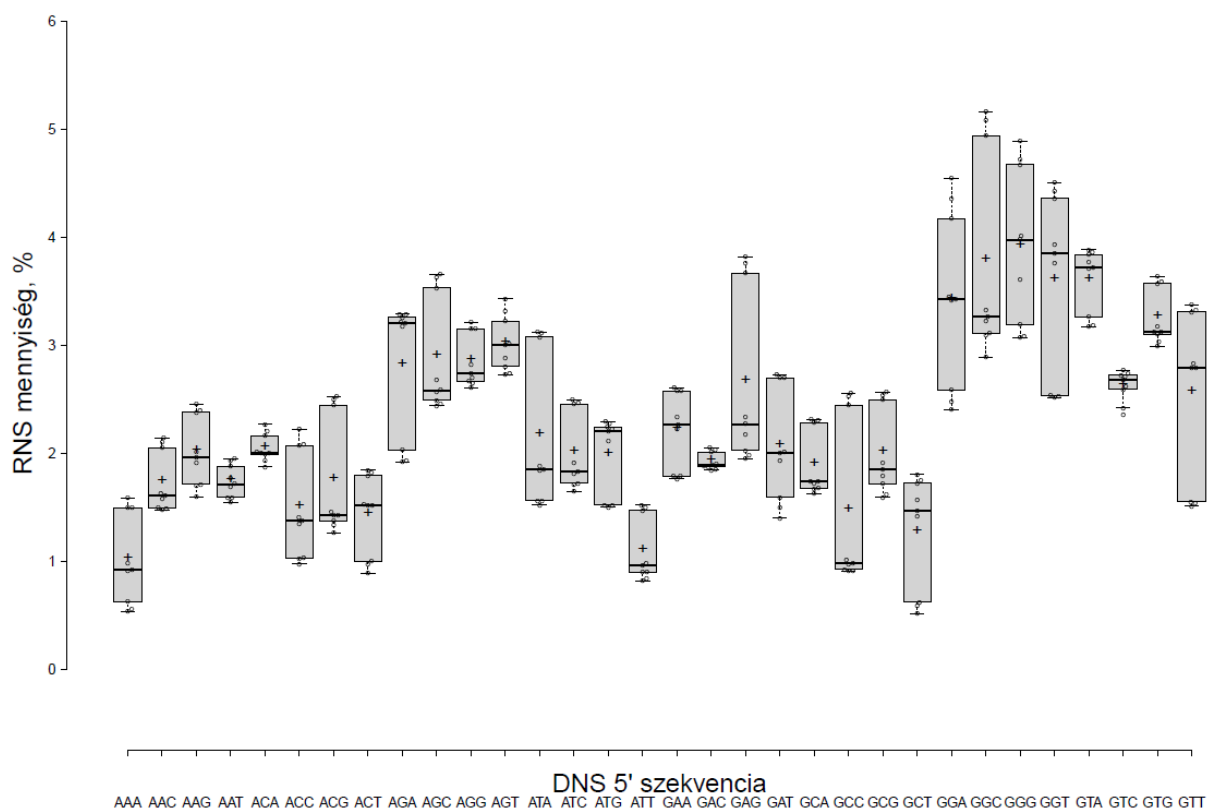


### 33. Ábra. A promótertől downstream elhelyezkedő első két nukleotid hatása az RNS expresszióra

Az ábrán a promótert követő első két nukleotid hatása látható a 7sk promóterről termelődött RNS-ek mennyiségére. Az ábra az összes transzkripció iniciációs pozíció transzkriptumainak mennyiségeit mutatja. 1 adatpont megadja az összes szekvenciáról termelődött RNS-ek összegéhez képest keletkezett RNS-ek százalékos arányát az adott szekvenciáról, adott cDNS könyvtárban belül. A boxokon belüli vastag vízszintes vonalak a mediánokat mutatják; a boxok határolják az R programmal meghatározott 25. és 75. percentilist; a szimpla vízszintes vonalak az interkvartilis tartomány 1,5-szereséig terjednek ki a 25. és 75. percentilistől; a keresztek a minták átlagát mutatják; az adatpontok körökként vannak ábrázolva. N=9, amely a 9 cDNS könyvtárnak felel meg. A p értékek a Függelékben a 112. oldalon találhatóak meg.

Mivel az eddig bemutatott eredmények alapján látható, hogy ha a +1-es pozícióban pirimidin van, a transzkripció iniciációs kezdő pozíciók sokkal heterogénebbek, ezért a további pozíciók szerepét az RNS mennyiségekre már csak +1A-val és +1 G-vel mutatom be a továbbiakban.

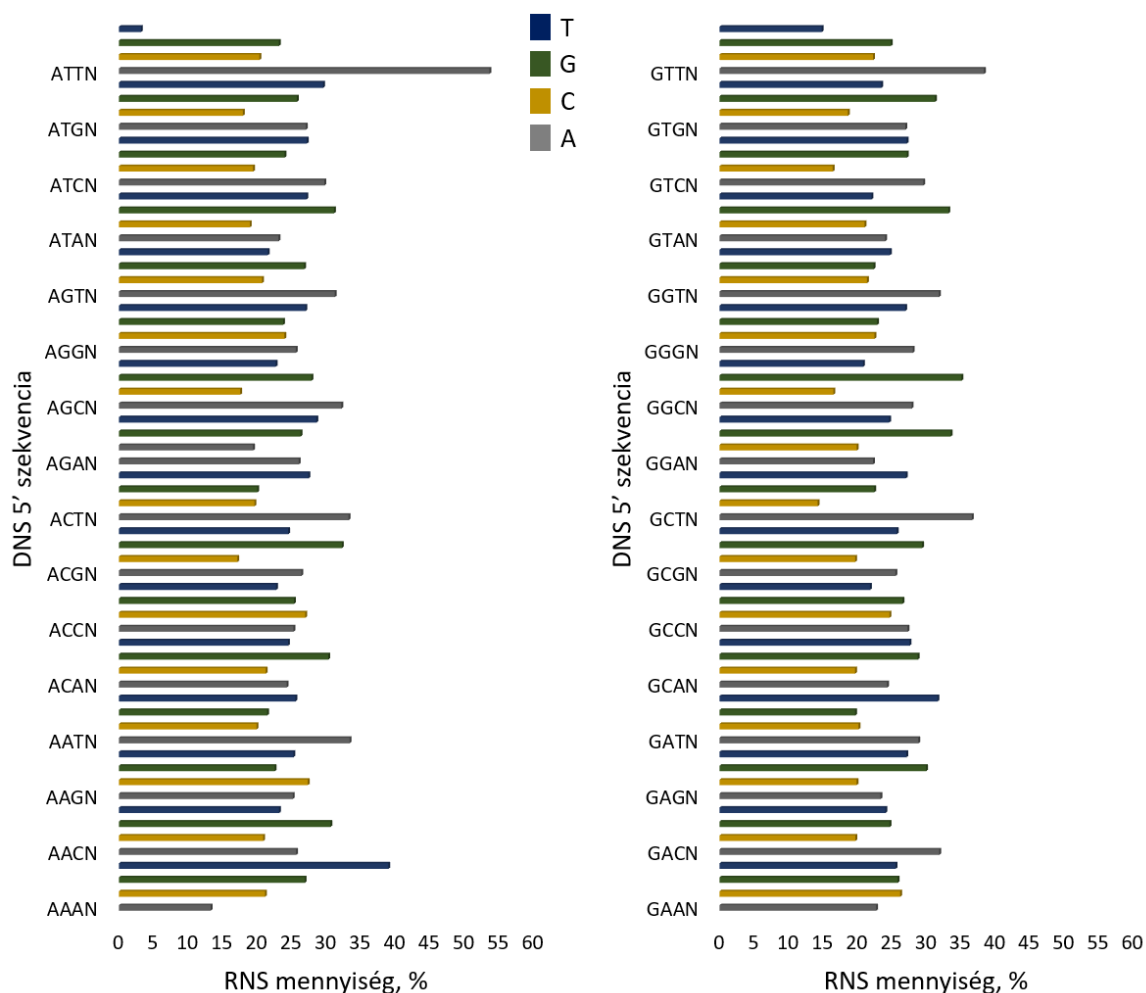
Ha a promótertől downstream elhelyezkedő első 3 pozíció szekvenciáját vizsgáltuk, azt kaptuk, hogy +1 A-val a legtöbb különbség az AG kezdetű dinukleotidok és a többi +1A kezdetű variáció között van. Az első 3 pozíció szekvenciáját vizsgálva +1 G-vel azt találtuk, hogy a +3-as pozícióban elhelyezkedő nukleotid is képes szignifikáns különbségeket okozni az RNS mennyiségben. A legtöbb szignifikáns különbség a GG és GT kezdetű dinukleotidok és a többiek között van (34. ábra). Eredményeink alapján a 7sk promóternél ha a promótertől downstream az első 2 pozícióban ugyanaz a dinukleotid van, viszont az azt követő pozícióban más nukleotid van, szignifikáns különbségeket képes okozni az RNS mennyiségekben. Eredményeink alapján a +1-es pozícióban a G egy kicsivel jobb RNS expresszióhoz vezethet, mint az A, de a további szekvencia függvényében +1A-ról is keletkezhet ugyanannyi, vagy több RNS, mint +1G-ről (lásd 33. és a 34. ábrák).



**34. Ábra.** A promótertől downstream elhelyezkedő első három nukleotid hatása az RNS expresszióra, ha a +1-es pozícióban A vagy G van

Az ábrán a promóterterttől downstream elhelyezkedő első három pozícióban lévő nukleotidok hatása látható a 7sk promóterről termelődött RNS-ek mennyiségére, ha az első pozícióban A vagy G van. Az ábra az összes transzkripció iniciációs pozíció transzkriptumainak mennyiségeit mutatja. 1 adatpont megadja az adott szekvenciáról keletkezett RNS-ek százalékos arányát az összes szekvenciáról termelődött RNS-ek összegéhez képest az adott cDNS könyvtárban. A boxokon belüli vastag vízszintes vonalak a mediánokat mutatják; a boxok határolják az R programmal meghatározott 25. és 75. percentilist; a szimpla vízszintes vonalak az interkvartilis tartomány 1,5-szereséig terjednek ki a 25. és 75. percentilistől; a keresztek a minták átlagát mutatják; az adatpontok körökként vannak ábrázolva. N=9, amely a 9 cDNS könyvtárnak felel meg. A p értékek a Függelékben a 113-114. oldalon találhatóak meg.

A +4-es pozíció hatását is megvizsgáltuk az RNS expresszióra, és azt találtuk, hogy az U6 promóterhez hasonlóan akár néhányszoros RNS mennyiségbeli különbségeket is képes okozni az ebben a pozícióban elhelyezkedő bázis is, ha a tőle upstream elhelyezkedő 3 bázis megegyezik. A 35. ábrán látható, hogy a +4-es pozíció az előtte lévő szekvencia függvényében változó mértékben befolyásolja az expressziós szintet.

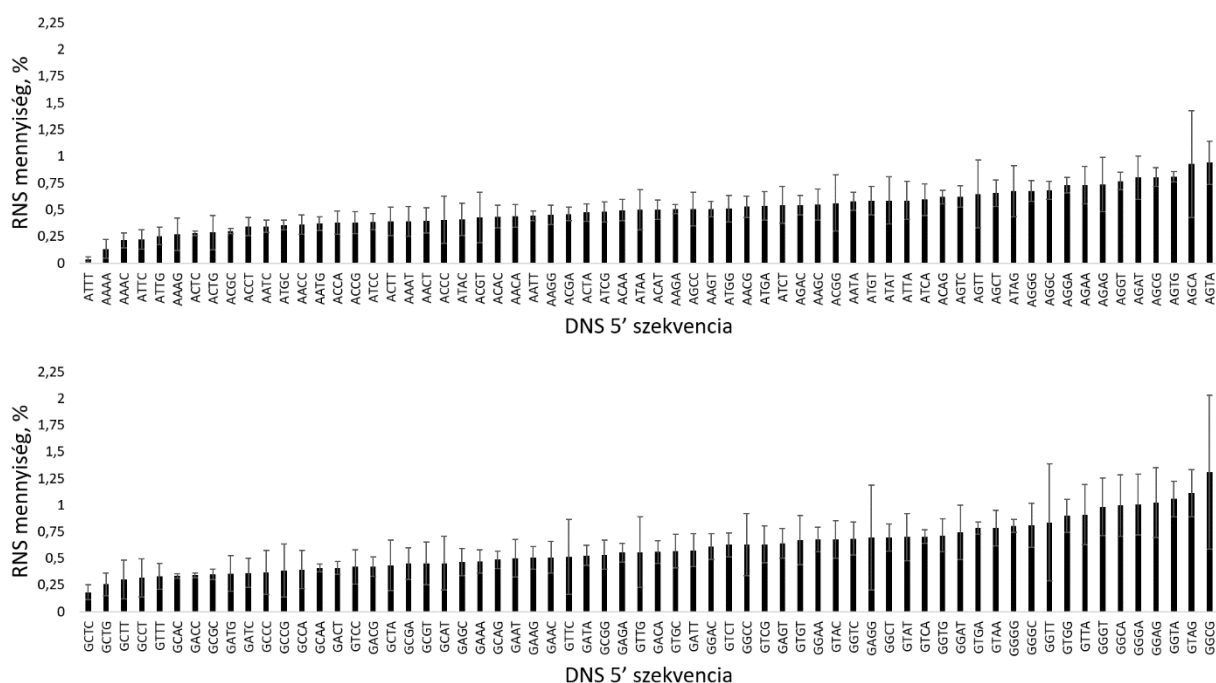


### 35. Ábra. A +4-es pozícióban lévő nukleotidok hatása az RNS expresszióra, ha a +1 A vagy G

Az ábrán a +4-es pozícióban elhelyezkedő különböző nukleotidok hatása látható a 7sk promóterről termelődött RNS-ek mennyiségére, ha a +1-es pozícióban A vagy G van. Az ábra az összes transzkripció iniciációs pozíció transzkriptumainak mennyiségeit mutatja. 1 oszlop megadja, hogy a promóter követő 4 bázis adott szekvenciájáról, 9 db cDNS könyvtár alapján átlagosan hány % RNS keletkezett az összes, első 3 pozíció szekvenciájában megegyező, de 4. pozícióban más nukleotidot tartalmazó szekvenciáról termelődött RNS-ek összegéhez képest.

A promótert downstream követő első 4 nukleotid szekvenciájának függvényében a fenti eredményeken látható módon változó mennyiségű RNS expresszálódik a 7sk promoterről. A legkisebb és a legnagyobb expressziót biztosító szekvenciákról termelődött RNS mennyiségek átlagai között közel 27x-es különbség van +1A-val (ATTT és AGTA) és több, mint 7x-es +1G-vel (GCTC és GGCG) (36. ábra).

Eredményeink alapján a 7sk promoternél nemcsak a +1-es pozíció, hanem a továbbiak is befolyásolják a keletkező RNS-ek mennyiségét.

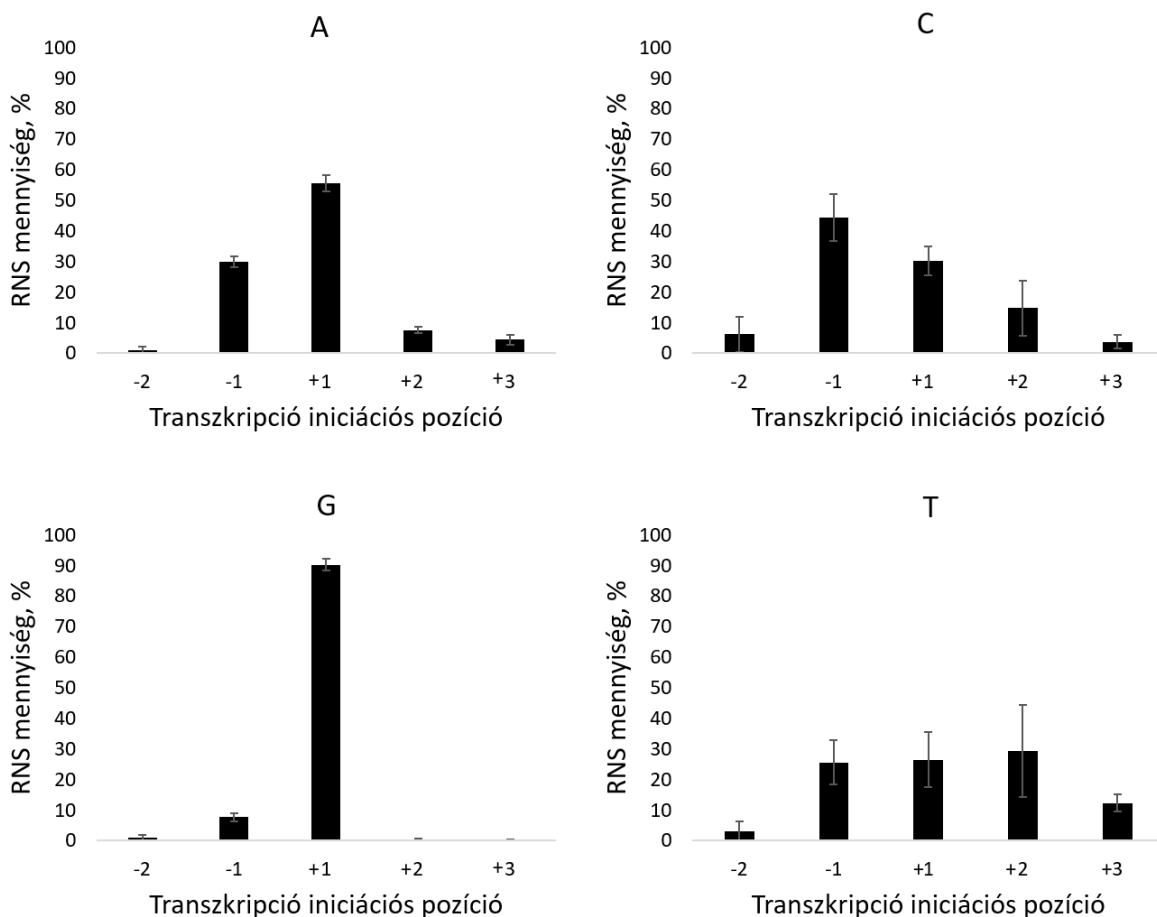


**36. Ábra.** A 7sk promotert downstream követő négy bázisos szekvencia hatása az RNS expresszióra, ha a +1 A vagy G

Az ábrán a +1-től a +4-es pozícióig terjedő szekvencia hatása látható a 7sk promoterről termelődött RNS-ek mennyiségére, ha az első pozícióban A, vagy G van. Az ábra az összes transzkripciós iniciációs pozíció transzkriptumainak mennyiségeit mutatja. 1 oszlop megadja, hogy az összes lehetséges szekvenciáról (256 variáció) termelődött RNS-ek összegéhez képest adott szekvenciáról az RNS-ek hány %-a keletkezett átlagosan. A hibásávok a százalékos értékek átlagainak szórását mutatják. N=9, amely a 9 cDNS könyvtárnak felel meg.

### 5.3.4.3. A T7 promóter

Az NGS eredményeink alapján először megvizsgáltuk a +1-es pozícióban lévő nukleotid szerepét a transzkripciós iniciációs kezdő pozíciókra (37. ábra). Ha a +1-ben A volt, akkor az összes RNS átlag 56%-a keletkezett a +1 pozícióról, +1G-vel ez 90% volt. Rendkívül kis arányban indul el pontosan az átírás, a purinokhoz képest, ha a +1 pirimidin, +1C-vel 30%, +1T-vel 26% volt a +1-ről átíródott RNS-ek aránya.

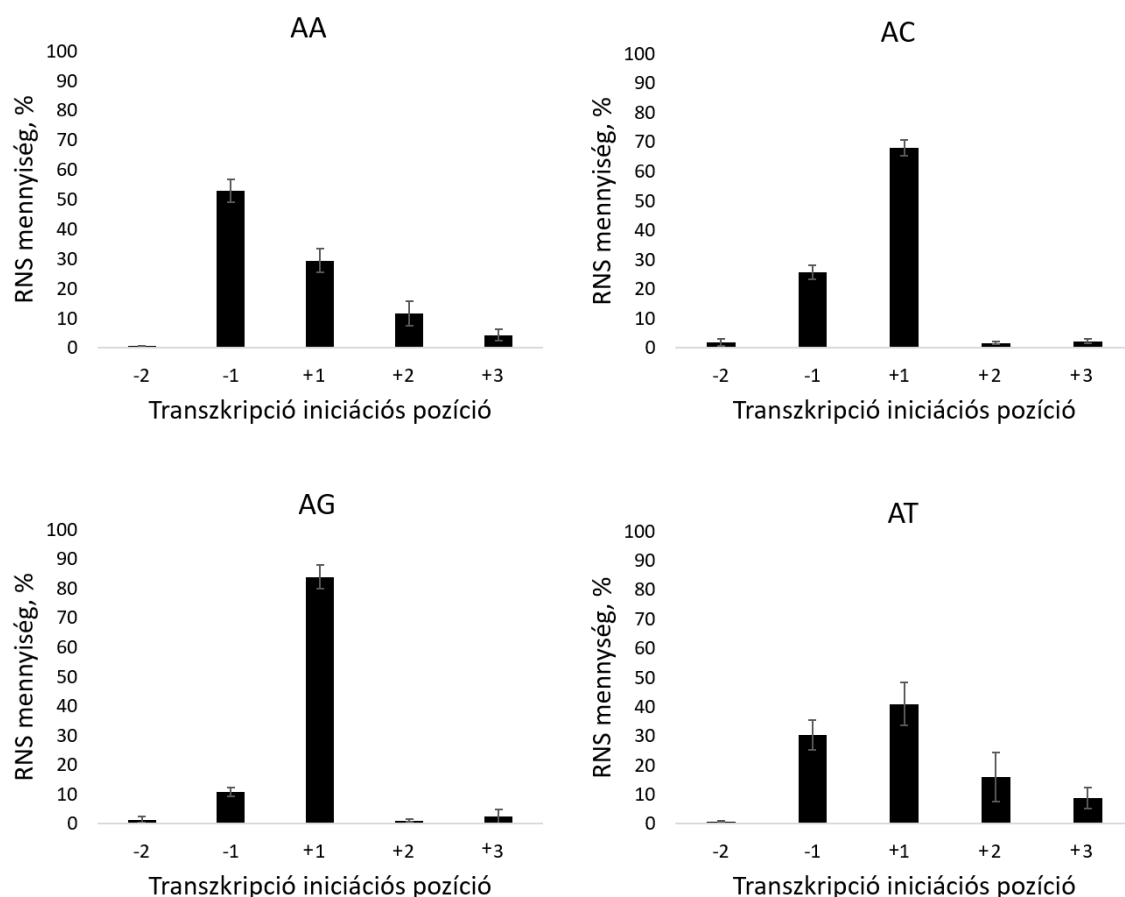


**37. Ábra. A T7 promótertől downstream elhelyezkedő első nukleotid hatása a különböző transzkripció iniciációs pozíciókra**

Az ábrán látható, hogy a T7 promótertől downstream elhelyezkedő első nukleotid miképp befolyásolja, hogy melyik pozícióról milyen arányban indul el az átírás. Egy oszlop megmutatja, hogy az összes vizsgált pozícióról termelődött RNS-ek összegének átlagban hány %-a íródott át adott pozícióról. A hibasávok a százalékos értékek átlagainak szórását mutatják. N=9, amely a 9 cDNS könyvtárnak felel meg.

Ezután megvizsgáltuk, hogy a további pozíciókban lévő nukleotidok hogyan befolyásolják a különböző transzkripció kezdő pozíciókat. Azt találtuk, ha a +1-es pozícióban guanin van, a további pozíciók már nem befolyásolják olyan nagymértékben a kezdő pozíciót, az RNS-ek legnagyobb részét +1-ről íródna át (38-40. ábra). Ha a promótertől downstream elhelyezkedő első pozícióban pirimidin van, akkor nagyobb arányban indul el vagy a -1-ről, vagy a +2-ről a transzkripció, mint a +1-ről, és számítanak a további pozíciók is abban, hogy melyik pozícióról milyen arányban indul el az átírás (39. ábra).

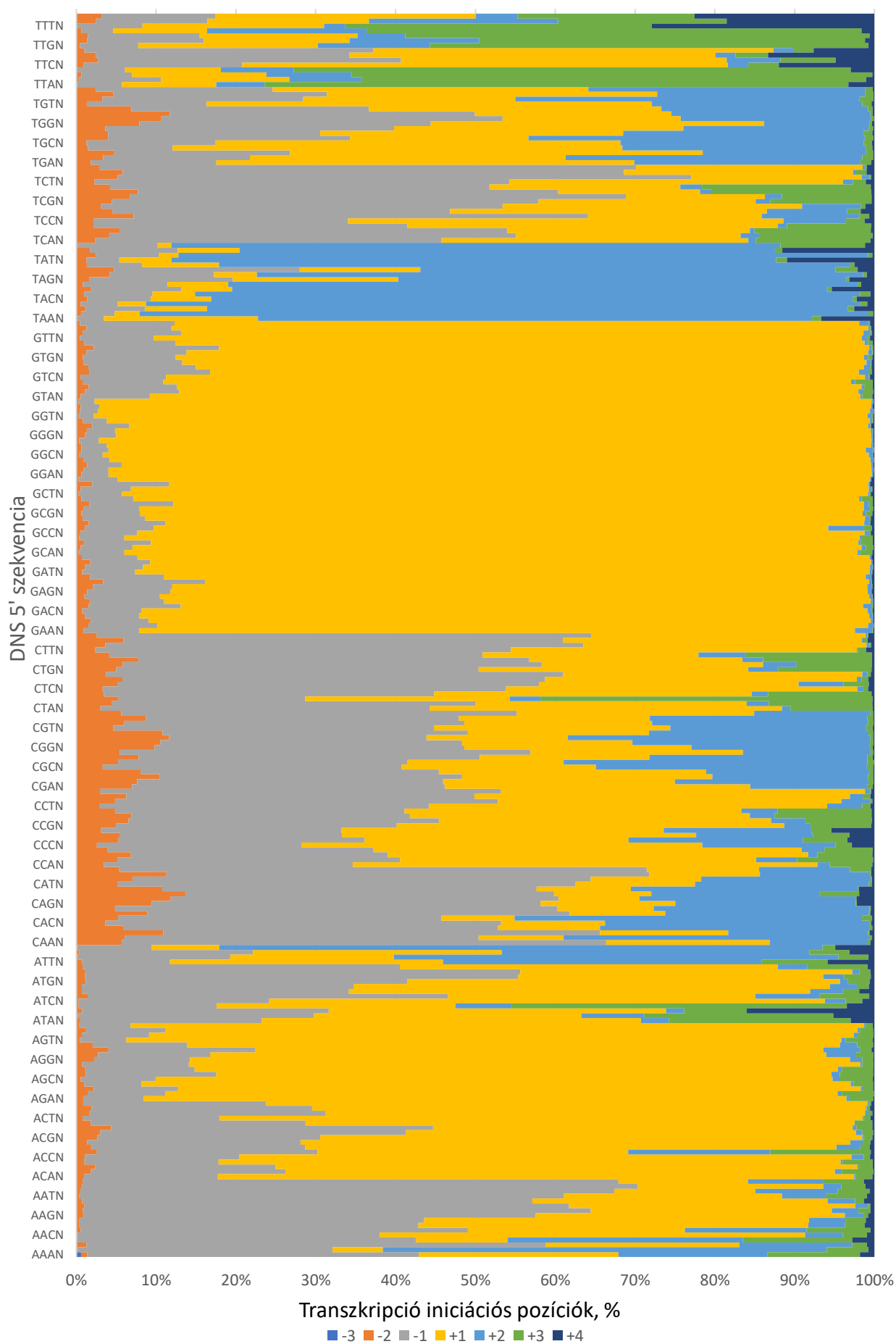




**38. Ábra. A +2-es pozíció nukleotidjának hatása a különböző transzkripció iniciációs pozíciókra, ha a +1 A**

Az ábrán látható, hogy a +2-es pozíció nukleotidja miképp befolyásolja, hogy melyik pozícióról milyen arányban indul el az átírás, ha a +1 A. Egy oszlop megmutatja, hogy az összes vizsgált pozícióról termelődött RNS-ek összegének átlagban hány %-a íródott át adott pozícióról. A hibásávok a százalékos értékek átlagának szórását mutatják. N=9, amely a 9 cDNS könyvtárnak felel meg.

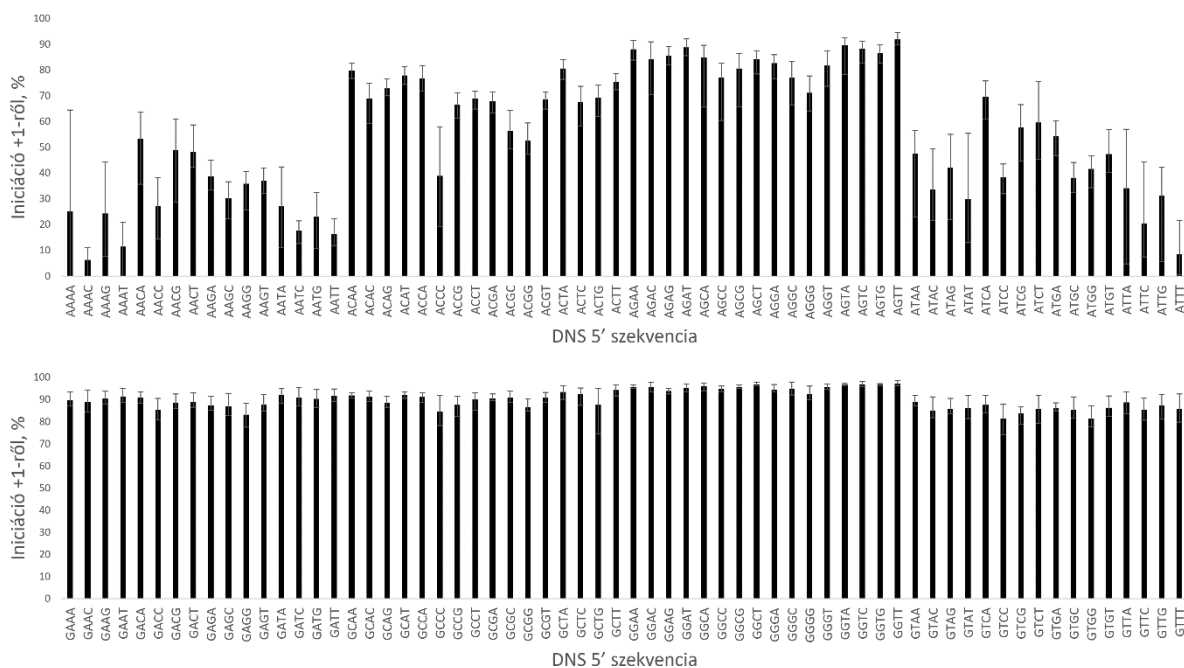
Míg +1G-vel szinte csak jó hosszúságú RNS-ek keletkeztek, legnagyobb arányban a +1-ben kezdődik az átírás, addig +1A-val a további pozíciók is eltéréseket okoztak (39-40. ábra). Ha a promótertől downstream az első nukleotid A, akkor a 2. pozícióban lévő nukleotid is nagy szerepet játszik a különböző transzkripciós iniciációs kezdőhelyekre: míg AA-ról csak átlagosan 29, AT-ről 41, addig AC-ről 68, AG-ről pedig 84 %-ban a +1-ről indul el az átírás (38. ábra). A +1-ben A-val a 39. és a 40. ábrákon látható, hogy a további pozíciók is szerepet játszanak abban, hogy melyik pozícióról milyen arányban indul el az átírás, hiába azonos az első 2 vagy 3 promótertől downstream elhelyezkedő pozíciók szekvenciája, de ha a +3, vagy +4-es pozícióban különböző nukleotidok vannak, akkor jelentősen módosulhat a +1-ről átíródó RNS-ek százalékos aránya.



### 39. Ábra. A T7 promótertől downstream elhelyezkedő szekvenciák hatása a különböző transzkripció iniciációs pozíciókra

Az ábrán látható, hogy a T7 promótertől downstream elhelyezkedő szekvenciák miképp befolyásolják, hogy melyik pozícióról milyen arányban indul el az átírás. Egy oszlop egy színe megmutatja, hogy az összes vizsgált pozícióról termelődött RNS-ek összegének átlagban hány %-a íródott át adott pozícióról, és egy oszlop egy darab kezdő szekvenciát jelöl. A közepesen világoskék szín a -3, a narancssárga szín a -2, a szürke szín a -1, a citromsárga szín a +1, a világoskék szín a +2, a zöld szín a +3 és a sötétkék szín a +4 pozíciókat jelöli a T7 promótertől downstream. N=9, amely a 9 cDNS könyvtárnak felel meg.

Azt találtuk, hogy a +1-es pozícióról induló, pontos transzkripció az AG kezdő dinukleotidokról is elérhető, átlagosan 84%-ban, amely majdnem eléri a +1-ben G-t tartalmazó szekvenciák pontosságát, az átlag 90%-ot (37. és 38. ábra). Eredményeink alapján az AG kezdetű szekvenciákról a transzkripció iniciációs pozíciót a +3-4. pozíciók szekvenciája már nem befolyásolja jelentősen, a +1G kezdetű szekvenciákhoz hasonlóan (40. ábra).

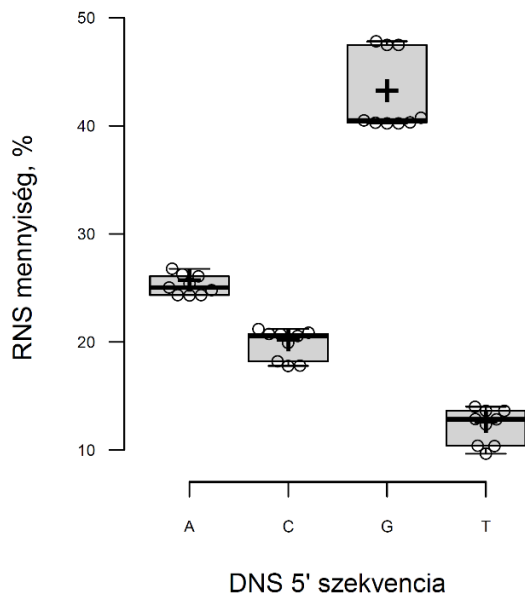


### 40. Ábra. A T7 promótertől downstream lévő szekvenciák hatása a +1-ről induló transzkripcióra, ha a +1 A, vagy G

Az ábrán a T7 promóter után elhelyezkedő különböző szekvenciák szerepe látható a +1 pozícióról történő átírásra. Az oszlopok megmutatják, hogy átlagban hány %-ban indult el a +1-ről az átírás adott templát DNS szekvencia esetén. A hibaszávok a százalékos értékek átlagainak szórását mutatják. N=9, amely a 9 cDNS könyvtárnak felel meg.

A következő lépésekben azt néztük meg, hogy a T7 promótertől downstream elhelyezkedő nukleotidok hogyan hatnak a transzkriptumok mennyiségére. A +1-es pozíció szerepét vizsgálva az átíródott RNS-ek mennyiségére azt találtuk, hogy ha a promóter után az első bázis purin, több RNS termelődik, mintha az első bázis pirimidin lenne.

Az első pozíciót vizsgálva a különböző kezdő nukleotidokról keletkezett transzkriptumok mennyisége egymástól szignifikánsan eltér, és a legtöbb RNS +1G-ről jött létre (41. ábra).

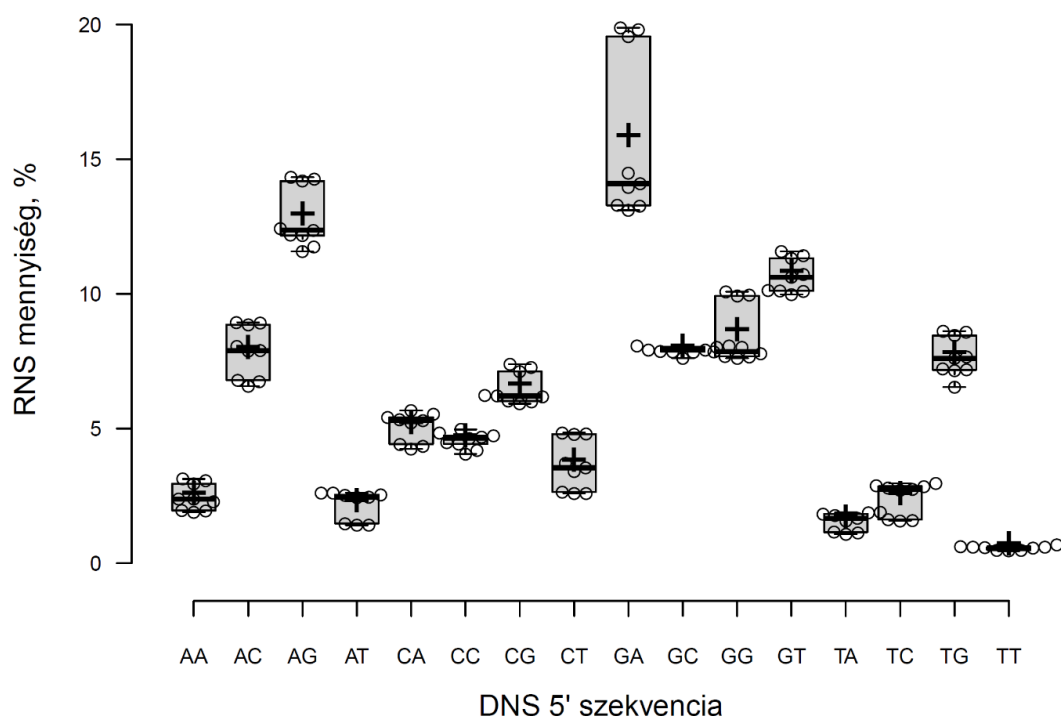


**41. Ábra. A +1-es pozícióban lévő nukleotid hatása az RNS expresszióra**

Az ábrán a +1-es pozícióban lévő nukleotid hatása látható a T7 promóterről termelődött RNS-ek mennyiségére. Az ábra az összes transzkripció iniciációs pozíció transzkriptumainak mennyiségeit mutatja. 1 adatpont megadja, hogy adott nukleotiddal az összes RNS hány %-a keletkezett az adott cDNS könyvtáron belül. A boxokon belüli vastag vízszintes vonalak a mediánokat mutatják; a boxok határolják az R programmal meghatározott 25. és 75. percentilist; a szimpla vízszintes vonalak az interkvartilis tartomány 1,5-szereséig terjednek ki a 25. és 75. percentilistől; a keresztek a minták átlagát mutatják; az adatpontok körökként vannak ábrázolva. N=9, amely a 9 cDNS könyvtárnak felel meg. A p értékek a Függelékben a 115. oldalon találhatók meg.

A promótert downstream követő első 2 nukleotidot vizsgálva azt találtuk, hogy a T7 promóternél is rendkívül fontos az átíródott RNS-ek mennyiségére nézve a +2-es pozíció is (42. ábra). Eredményeink alapján a legtöbb RNS GA és meglepő módon AG kezdettel keletkezik. AG és GA szignifikánsan a legjobbak a 16 variáció közül és az ezekről a kezdő szekvenciákról keletkezett RNS-ek mennyisége egymástól szignifikánsan nem különbözik. A GT és a GA is szignifikánsan több RNS-t eredményeztek, mint a GC és a GG. Eredményeink alapján, ha a promóter utáni első két pozíció szekvenciáját vizsgáljuk, nem keletkezik több RNS GG-vel, mint ha a +1 G-t másik nukleotid követi, sőt a GA kezdő szekvenciáról átlagosan mintegy kétszer annyi RNS expresszálódik, mint GG-ről.

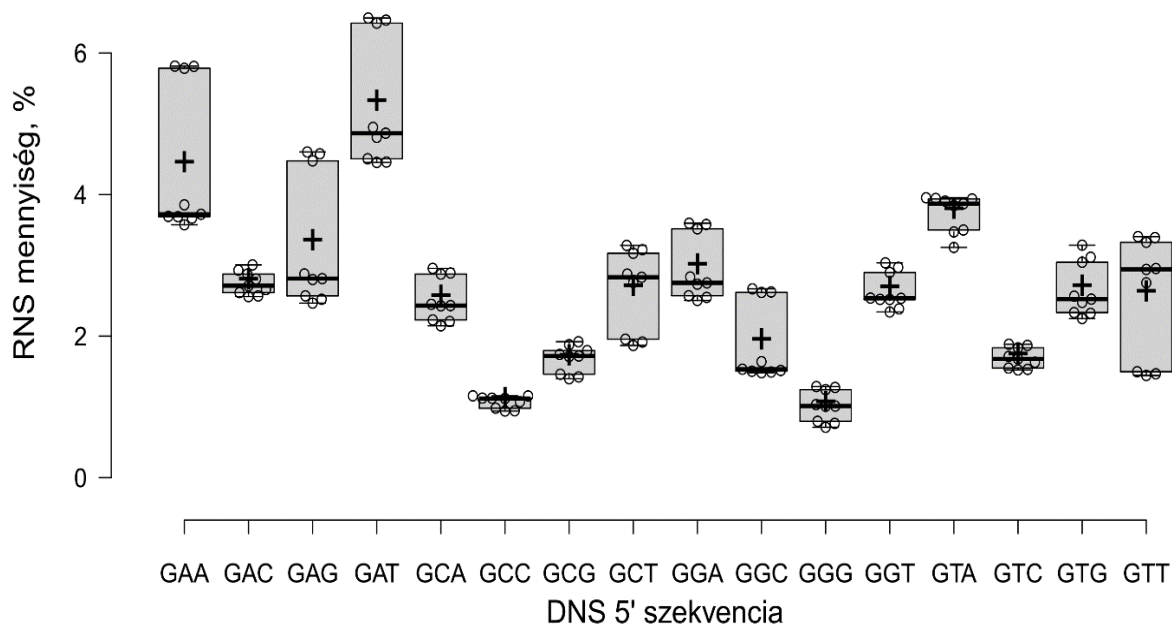
A 42. ábrán látható eredményeink alapján keletkezhet a T7 promóterről olyan mennyiségű RNS +1A-val, sőt a +1-es pozícióban pirimidinnel is, mint +1G-vel, és +1 G-vel is különbség van az RNS szintekben a +2-ben elhelyezkedő nukleotidtól függően.



#### 42. Ábra. A promótertől downstream elhelyezkedő első két nukleotid hatása az RNS expresszióra

Az ábrán a promótert követő első két nukleotid hatása látható a T7 promóterről termelődött RNS-ek mennyiségére. Az ábra az összes transzkripciós iniciációs pozíció transzkriptumainak mennyiségeit mutatja. 1 adatpont megadja az összes szekvenciáról termelődött RNS-ek összegéhez képest keletkezett RNS-ek százalékos arányát az adott szekvenciáról, adott cDNS könyvtárban belül. A boxokon belüli vastag vízszintes vonalak a mediánokat mutatják; a boxok határolják az R programmal meghatározott 25. és 75. percentilist; a szimpla vízszintes vonalak az interkvartilis tartomány 1,5-szereséig terjednek ki a 25. és 75. percentilistől; a keresztek a minták átlagát mutatják; az adatpontok körökként vannak ábrázolva. N=9, amely a 9 cDNS könyvtárnak felel meg. A p értékek a Függelékben a 115. oldalon találhatók meg.

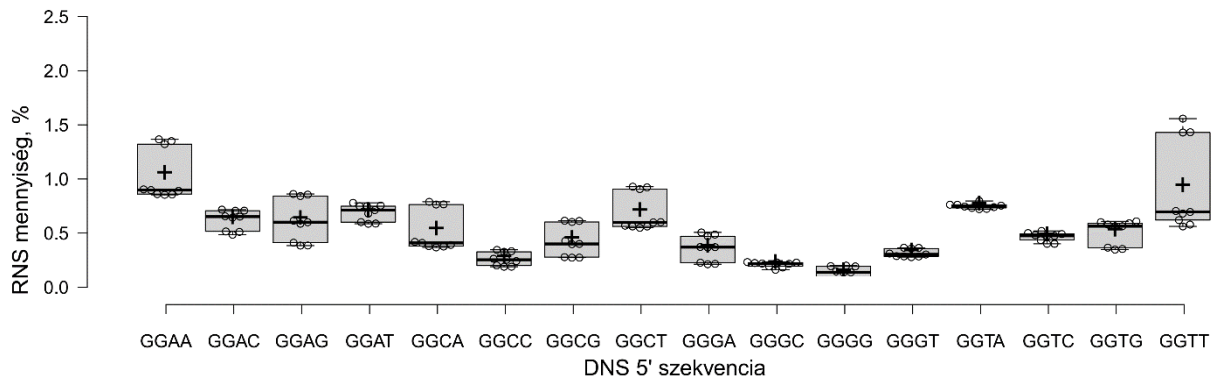
Mivel az eddig bemutatott eredményeink alapján, összességében akkor keletkezik a legtöbb RNS, ha a +1-ben G van, és ilyenkor az RNS-ek nagyrészt pontosan a +1-ről íródnak át, ezért a további pozíciók szerepét az RNS mennyiségekre már csak +1 G-vel mutatom be a továbbiakban. Ezek közül az első 3 pozíció szekvenciáját vizsgálva azt találtuk, hogy összességében a GA kezdetű szekvenciák vezettek a legtöbb RNS-hez és a GGG szekvencia a promótertől downstream a legalacsonyabb RNS szintet eredményező +1-ben G-t tartalmazó szekvenciák közé tartozik. Eredményeink alapján a T7 promóternél is fontos szerepet játszik a +3-as pozíció és ez is szignifikáns különbségeket okozhat az RNS mennyiségben (43. ábra).



**43. Ábra. A +2 és +3 pozíciókban lévő nukleotidok hatása az RNS expresszióra, ha a +1-es pozícióban G van**

Az ábrán a +2 és +3 pozíciókban lévő nukleotidok hatása látható a T7 promóterről termelődött RNS-ek mennyiségére, ha a +1-es pozícióban G van. Az ábra az összes transzkripciós iniciációs szekvencia transzkriptumainak mennyiségeit mutatja. 1 adatpont megadja az adott szekvenciáról keletkezett RNS-ek százalékos arányát az összes szekvenciáról termelődött RNS-ek összegéhez képest az adott cDNS könyvtárban. A boxokon belüli vastag vízszintes vonalak a mediánokat mutatják; a boxok határolják az R programmal meghatározott 25. és 75. percentilist; a szimpla vízszintes vonalak az interkvartilis tartomány 1,5-szereséig terjednek ki a 25. és 75. percentilistől; a keresztek a minták átlagát mutatják; az adatpontok körökként vannak ábrázolva. N=9, amely a 9 cDNS könyvtárnak felel meg. A p értékek a Függelékben a 116. oldalon találhatók meg.

A promótertől downstream 4. pozícióban lévő nukleotid hatását a T7 promóterről átíródott transzkriptumok mennyiségére a templátban GG-vel kezdődő szekvenciákkal mutatom be. Azt találtuk, hogy még a 4. pozíció is egyes esetekben okozott szignifikáns különbséget az RNS mennyiségben és a GGGG kezdetű szekvenciáról a GGGC kivételével az összes vizsgált szekvenciához képest szignifikánsan a legkevesebb RNS keletkezett (44. ábra). Eredményünk alapján akár néhányszoros RNS mennyiségbeli különbségeket is képes okozni a 4. pozícióban elhelyezkedő nukleotid, ha az első 3 megegyezik, tehát még a 4. pozícióban lévő nukleotid is befolyásolja a transzkripciót.



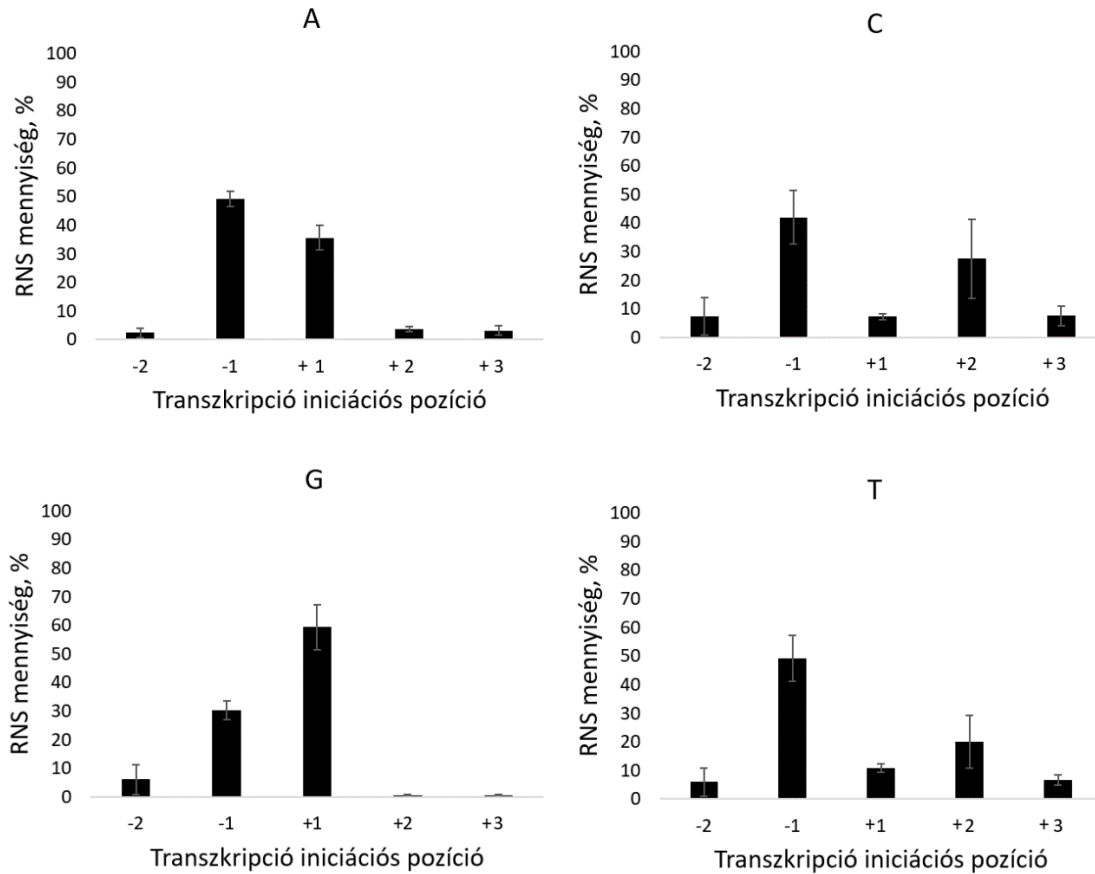
**44. Ábra. A +3 és +4 pozíciókban lévő nukleotidok hatása az RNS expresszióra, ha a promótertől downstream elhelyezkedő első két nukleotid GG**

Az ábrán a +3 és +4 pozíciókban lévő nukleotidok hatása látható a T7 promóterről termelődött RNS-ek mennyiségére, ha a promótertől downstream elhelyezkedő első két nukleotid GG. Az ábra az összes transzkripciós iniciációs pozíció transzkriptumainak mennyiségeit mutatja. 1 adatpont megadja az adott szekvenciáról keletkezett RNS-ek százalékos arányát az összes szekvenciáról termelődött RNS-ek összegéhez képest az adott cDNS könyvtárban. A boxokon belüli vastag vízszintes vonalak a mediánokat mutatják; a boxok határolják az R programmal meghatározott 25. és 75. percentilist; a szimpla vízszintes vonalak az interkvartilis tartomány 1,5-szereséig terjednek ki a 25. és 75. percentilistől; a keresztek a minták átlagát mutatják; az adatpontok körökként vannak ábrázolva. N=9, amely a 9 cDNS könyvtárnak felel meg. A p értékek a Függelékben a 116. oldalon találhatók meg.

A promótertől downstream elhelyezkedő első 4 bázispár hosszú szekvencia függvényében a fenti eredmények alapján látható módon változó mennyiségű RNS expresszálódik a T7 promóterről. A legkisebb és a legnagyobb expressziót biztosító szekvenciákról termelődött RNS mennyiségek átlagai között közel 50x-es különbség van +1A-val (AAAC és AGAA), közel 5x-ös AG kezdettel (AGGG és AGAA) és több, mint 13x-os +1G-vel (GGGG és GATA) (45. ábra). A 45. ábrán látható, hogy a legnagyobb expressziót biztosító 4 bázispáros szekvenciák +1A-val a 2. pozícióban G-t vagy C-t tartalmaznak.



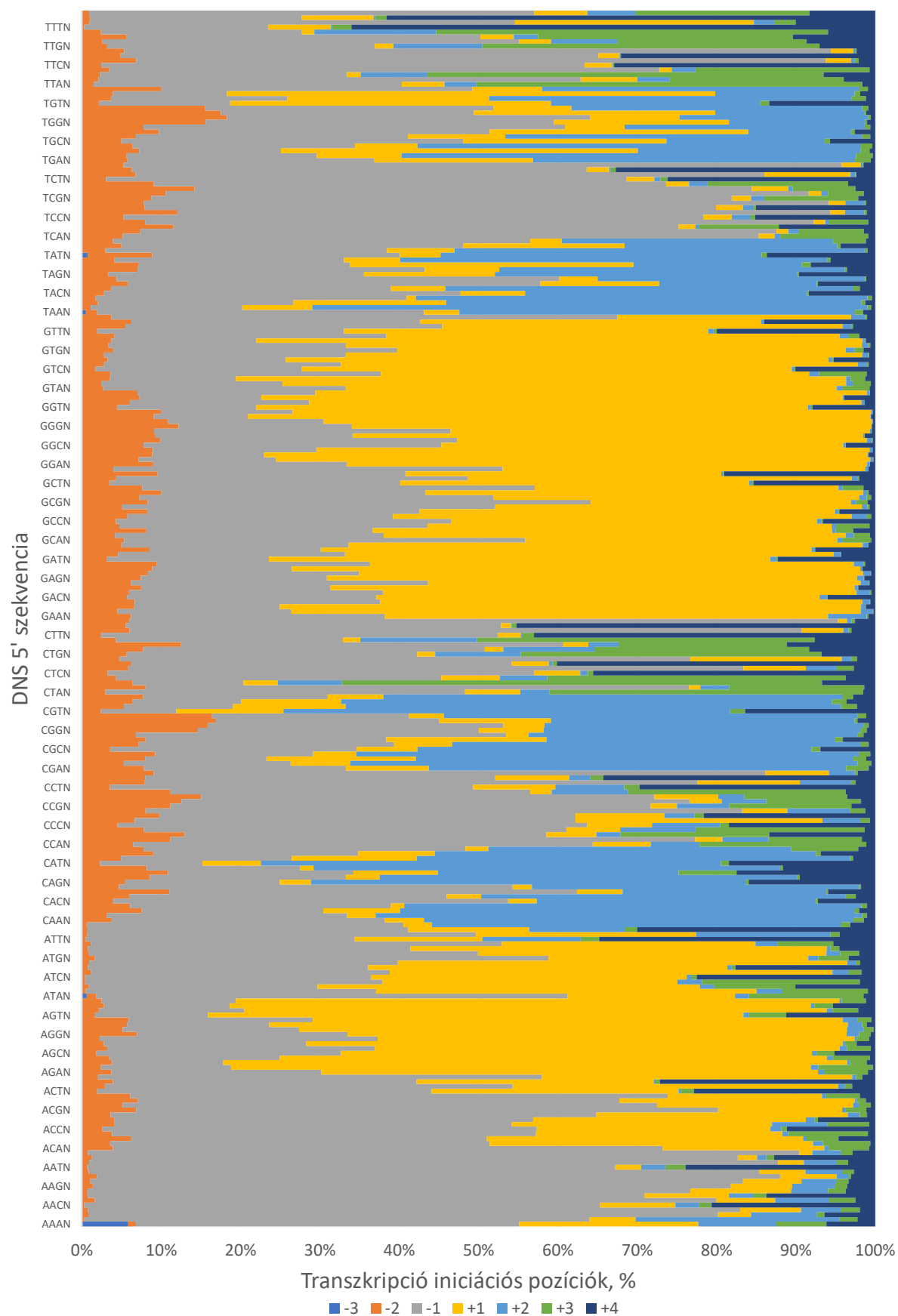




#### 46. Ábra. A +1-es pozícióban lévő nukleotid hatása a különböző transzkripció iniciációs pozíciókra

Az ábrán látható, hogy az U6T7 promótertől downstream elhelyezkedő első nukleotid miképp befolyásolja, hogy melyik pozícióról milyen arányban indul el az átírás. Egy oszlop megmutatja, hogy az összes vizsgált pozícióról termelődött RNS-ek összegének átlagban hány %-a íródott át adott pozícióról. A hibaszávok a százalékos értékek átlagainak szórását mutatják. N=9, amely a 9 cDNS könyvtárnak felel meg.

Ezután megvizsgáltuk a további pozíciókban lévő nukleotidok szerepét is a különböző transzkripció kezdő pozíciókra. A 47. ábra alapján látható, hogy a további pozícióknak ennél a promóternél is szerepe van a transzkripció iniciáció pozíciójának meghatározásában. +1G-vel indul el a legnagyobb arányban +1-ről az átírás, de a többi promóterhez képest ezt a +3-4-es pozíciók is jobban befolyásolják. +1A-ról, ha a +2 A, szinte alig keletkezik RNS a +1-es pozícióról, míg ha a +2 G, akkor a legtöbb (ha csak az első két pozíciót tekintjük) és a további pozíciók is befolyásolják, hogy milyen arányban indul el +1-ről az átírás (47. ábra).

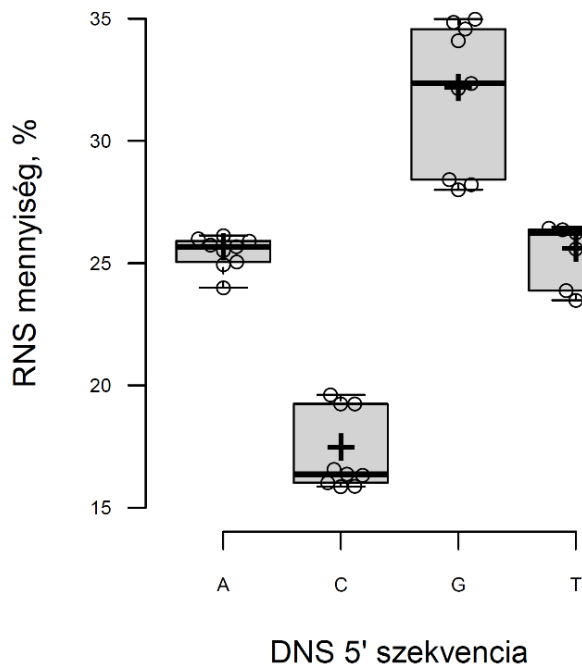


#### **47. Ábra. Az U6T7 promótertől downstream elhelyezkedő szekvenciák hatása a különböző transzkripció iniciációs pozíciókra**

Az ábrán látható, hogy az U6T7 promótertől downstream elhelyezkedő szekvenciák miképp befolyásolják, hogy melyik pozícióról milyen arányban indul el az átírás. Egy oszlop egy színe megmutatja, hogy az összes vizsgált pozícióról termelődött RNS-ek összegének átlagban hány %-a íródott át adott pozícióról, és egy oszlop egy darab kezdő szekvenciát jelöl. A közepesen világoskék szín a -3, a narancssárga szín a -2, a szürke szín a -1, a citromsárga szín a +1, a világoskék szín a +2, a zöld szín a +3 és a sötétkék szín a +4 pozíciókat jelöli az U6T7 promótertől downstream. N=9, amely a 9 cDNS könyvtárnak felel meg.

Ha a +1-ben pirimidin van csak elvétele a +1 a transzkripció kezdő pozíciója. Ha a +1-es pozícióban pirimidin van, a +2-ben purin, akkor nagyobb arányban indul el a +2-ről, vagy a -1-ről a transzkripció, mint a +1-ről. Ha a promótertől downstream elhelyezkedő első 2 pozícióban pirimidin van, de a harmadikban purin, szintén megfigyelhető, hogy megnő azoknak az RNS-eknek az aránya, amelyek a -1-es kezdő pozícióról keletkeztek (47. ábra). Eredményeink azt mutatják, hogy ha az U6T7 promóternél a +1-ben purin van, az átírás a többi vizsgált promóterrel szemben akkor sem csak a +1-ről indul el, míg ha a +1 pirimidin, az átírás nagyon ritkán történik +1-ről, és azt, hogy melyik pozícióról milyen arányban történik a transzkripció iniciációja, nagymértékben befolyásolják a promótertől downstream elhelyezkedő további nukleotidok is.

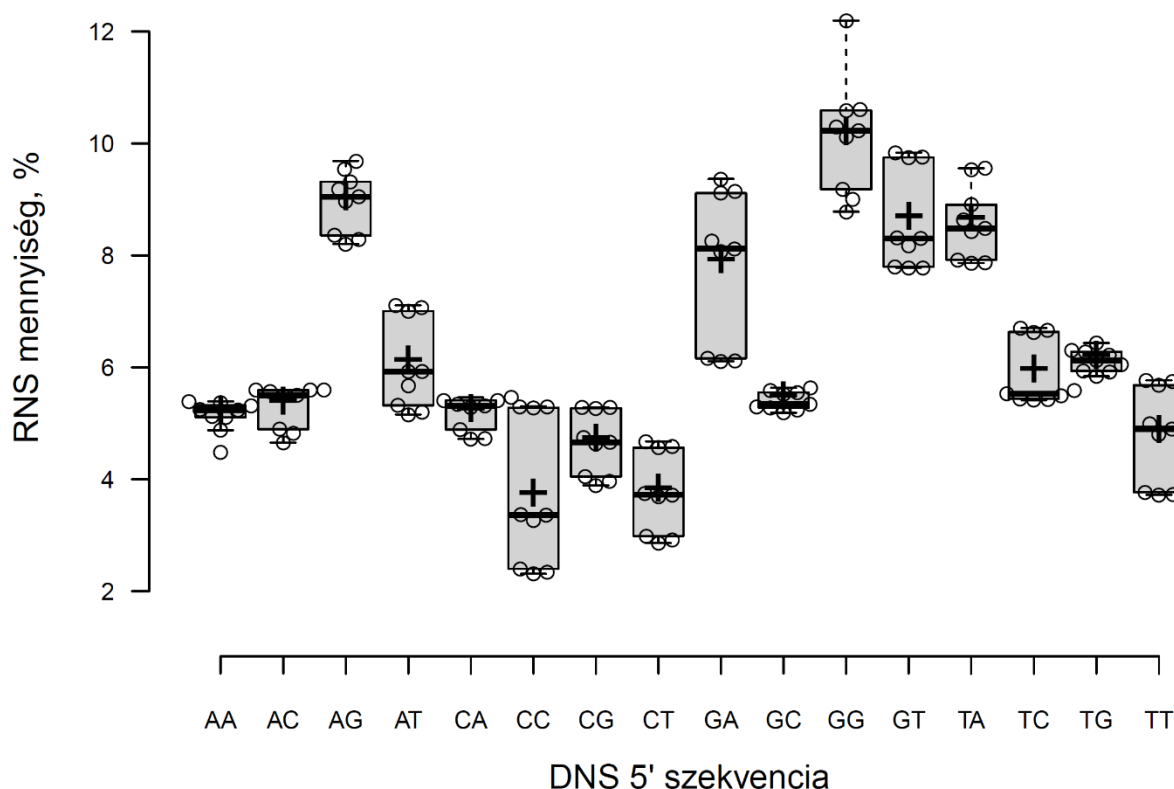
A következő lépésekben azt néztük meg, hogy az U6T7 promótertől downstream elhelyezkedő pozíciók hogyan hatnak a transzkriptumok mennyiségére (48. ábra). Az első nukleotid szerepét vizsgálva az átíródott RNS-ek mennyiségére azt találtuk, hogy nincsenek olyan nagy különbségek, mint a T7, U6, 7sk promótereknél. A +1C-ről szignifikánsan kevesebb RNS keletkezett, mint a többiről. A A és T között nem találtunk szignifikáns különbséget, a G pedig mindegyiknél szignifikánsan több RNS-t eredményezett



#### 48. Ábra. A +1-es pozícióban lévő nukleotid hatása az RNS expresszióra

Az ábrán a +1-es pozícióban lévő nukleotid hatása látható az U6T7 promóterről termelődött RNS-ek mennyiségére. Az ábra az összes transzkripciós iniciációs pozíció transzkriptumainak mennyiségeit mutatja. 1 adatpont megadja, hogy adott nukleotiddal az összes RNS hány %-a keletkezett az adott cDNS könyvtáron belül. A boxokon belüli vastag vízszintes vonalak a mediánokat mutatják; a boxok határolják az R programmal meghatározott 25. és 75. percentilist; a szimpla vízszintes vonalak az interkvartilis tartomány 1,5-szereséig terjednek ki a 25. és 75. percentilistól; a keresztet a minták átlagát mutatják; az adatpontok körökként vannak ábrázolva. N=9, amely a 9 cDNS könyvtárnak felel meg. A p értékek a Függelékben a 117. oldalon találhatók meg.

A promótert downstream követő első 2 pozíciót vizsgálva, azt kaptuk, hogy ennél a promóternél is befolyásolja a +2 is a promóterről keletkezett RNS-ek mennyiségét (49. ábra). Eredményeink alapján ennél a promóternél is keletkezhet olyan mennyiségű RNS a +1-es pozícióban pirimidinnel, mint purinnal, ha megfelelő nukleotid követi. Amiatt, hogy nem elég pontos a +1-ről történő átírás még akkor sem, ha a +1-ben purin van, az U6T7 promóterrel kapcsolatos további eredmények nem kerülnek bemutatásra.

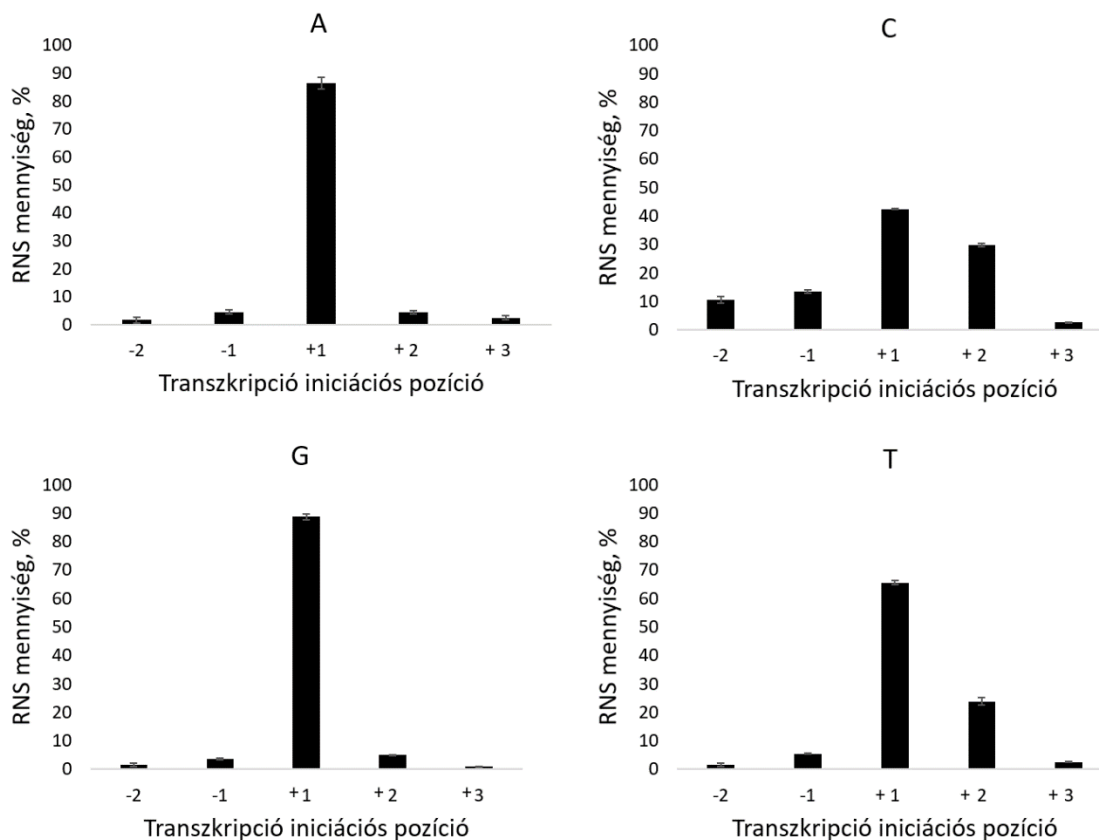


#### 49. Ábra. A promótertől downstream elhelyezkedő első két nukleotid hatása az RNS expresszióra

Az ábrán a promótert követő első két nukleotid hatása látható az U6T7 promóterről termelődött RNS-ek mennyiségére. Az ábra az összes transzkripciós iniciációs pozíció transzkriptumainak mennyiségeit mutatja. 1 adatpont megadja az összes szekvenciáról termelődött RNS-ek összegéhez képest keletkezett RNS-ek százalékos arányát az adott szekvenciáról, adott cDNS könyvtárban belül. A boxokon belüli vastag vízszintes vonalak a mediánokat mutatják; a boxok határolják az R programmal meghatározott 25. és 75. percentilist; a szimpla vízszintes vonalak az interkvartilis tartomány 1,5-szereséig terjednek ki a 25. és 75. percentilistől; a keresztek a minták átlagát mutatják; az adatpontok körökként vannak ábrázolva. N=9, amely a 9 cDNS könyvtárnak felel meg. A p értékek a Függelékben a 117. oldalon találhatóak meg.

#### 5.3.4.5. A J23119 promóter

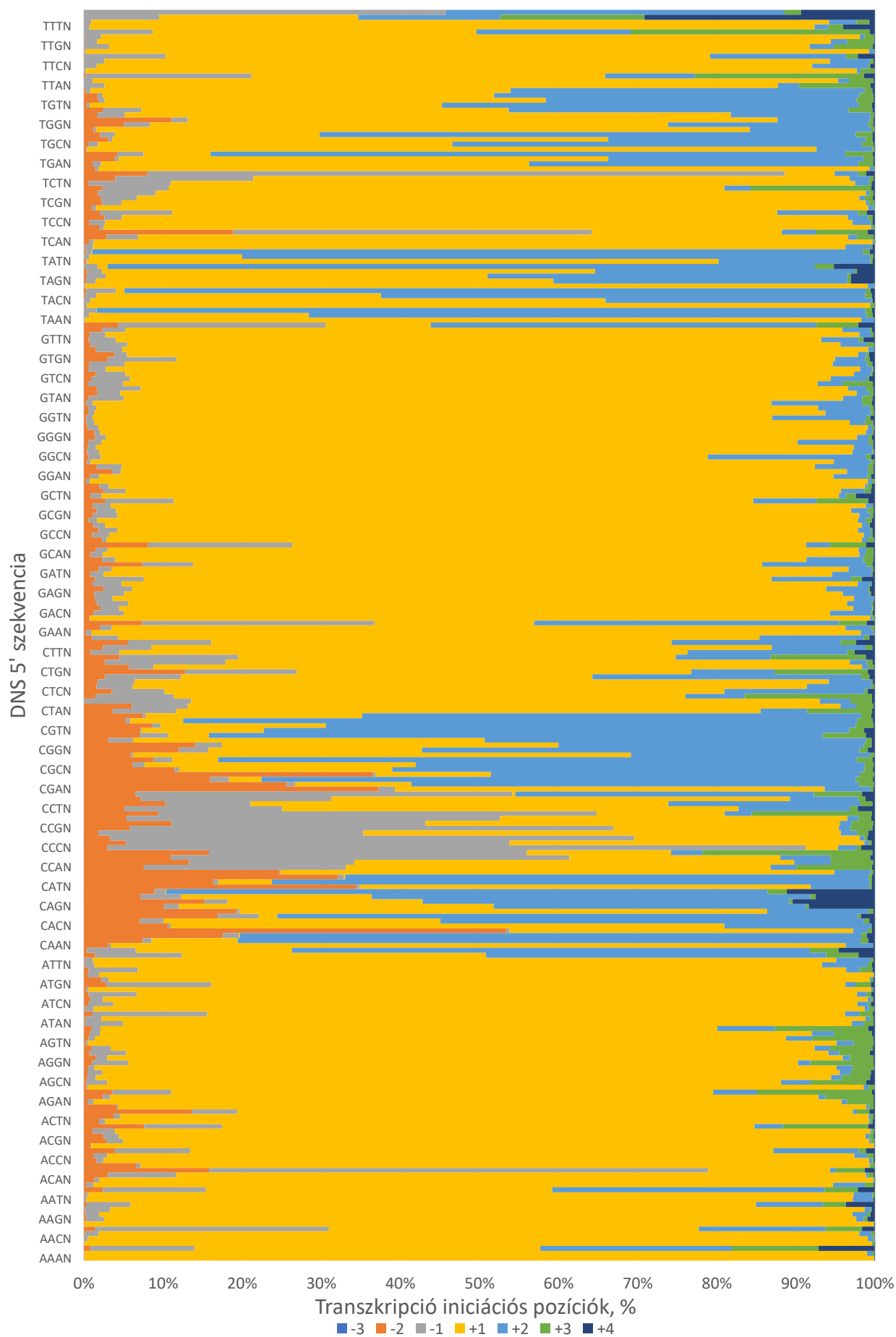
A J23119 promóterről eddig nem ismert, hogy a promótertől downstream elhelyezkedő nukleotidok befolyásolják-e a transzkripció iniciációjának pontos helyét. Az NGS eredményeink alapján először megvizsgáltuk +1-es pozíció szerepét a transzkripciós iniciációs kezdő pozíciókra (50. ábra). Ha a +1-ben A volt, akkor az összes RNS átlag 86%-a keletkezett a +1 pozícióról, +1G-vel pedig 89%. Habár a purinokhoz képest kisebb arányban indul el pontosan az átírás, ha a +1 pirimidin, de a többi vizsgált promóterhez képest több a +1-ből átíródó RNS pirimidinekről is. +1C-vel 42%, +1T-vel 65% volt a +1-ről átíródott RNS-ek aránya.



**50. Ábra. A J23119 promótertől downstream elhelyezkedő első nukleotid hatása a különböző transzkripció iniciációs pozíciókra**

Az ábrán látható, hogy a J23119 promótertől downstream elhelyezkedő első nukleotid miképp befolyásolja, hogy melyik pozícióról milyen arányban indul el az átírás. Egy oszlop megmutatja, hogy az összes vizsgált pozícióról termelődött RNS-ek összegének átlagban hány %-a íródott át adott pozícióról. A hibásávok a százalékos értékek átlagának szórását mutatják. N=3, amely a 3 cDNS könyvtárnak felel meg.

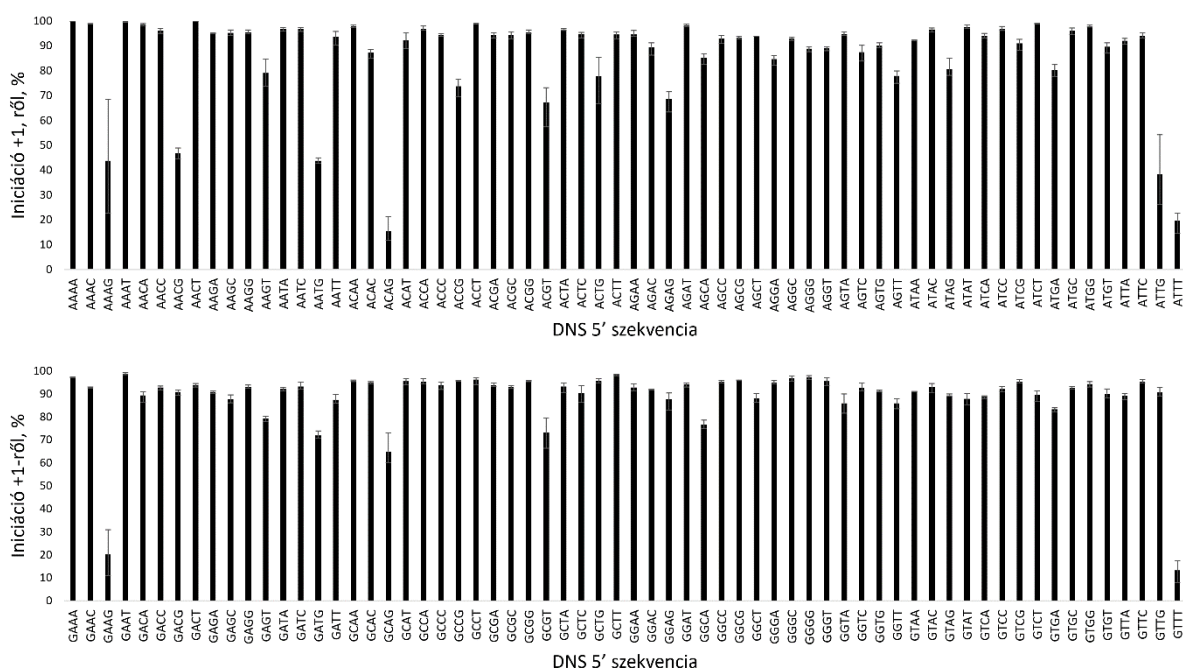
Ezután megvizsgáltuk a további pozíciókban lévő nukleotidok szerepét is a különböző transzkripció kezdő pozíciókra. Eredményeink azt mutatják, hogy ha a +1-es pozícióban purin van, az átírás majdnem mindig a +1-ről indul el, míg ha pirimidin, az átírás nemcsak +1-ről történik. Azt, hogy pirimidin esetében melyik pozícióról milyen arányban történik a transzkripció iniciációja, befolyásolják a promótertől downstream elhelyezkedő további nukleotidok is. Azt találtuk, hogy vannak olyan pirimidin kezdetű szekvenciák, amelyeknél megközelíti a +1-es pozícióból átíródó RNS-ek aránya a purin kezdetűekét, ilyenek a CT, TC, TT kezdetű szekvenciák. Ha a +1-ben pirimidin van, és a +2-ben purin, akkor nagyobb arányban indul el a +2-ről a transzkripció, mint a +1-ről (51. ábra).



### 51. Ábra. A J23119 promótertől downstream elhelyezkedő szekvenciák hatása a különböző transzkripció iniciációs pozíciókra

Az ábrán látható, hogy a J23119 promótertől downstream elhelyezkedő szekvenciák miképp befolyásolják, hogy melyik pozícióról milyen arányban indul el az átírás. Egy oszlop egy színe megmutatja, hogy az összes vizsgált pozícióról termelődött RNS-ek összegének átlagban hány %-a íródott át adott pozícióról, és egy oszlop egy darab kezdő szekvenciát jelöl. A közepesen világoskék szín a -3, a narancssárga szín a -2, a szürke szín a -1, a citromsárga szín a +1, a világoskék szín a +2, a zöld szín a +3 és a sötétkék szín a +4 pozíciókat jelöli a J23119 promótertől downstream. N=3, amely a 3 cDNS könyvtárnak felel meg.

Eredményeink alapján, ha az első pozícióban purin van, szinte csak jó hosszúságú RNS-ek keletkeznek, majdnem mindig a +1-es pozícióról kezdődik az átírás. Viszont van néhány szekvencia, +1A-ról kicsit több, mint +1G-ről, amelyiknél drasztikusan kevesebb jó hosszúságú RNS keletkezett (52. ábra). Azonban összességében nincs jelentős eltérés a +1-es pozícióban A és G között abban, hogy milyen pontos a transzkripció iniciációja.

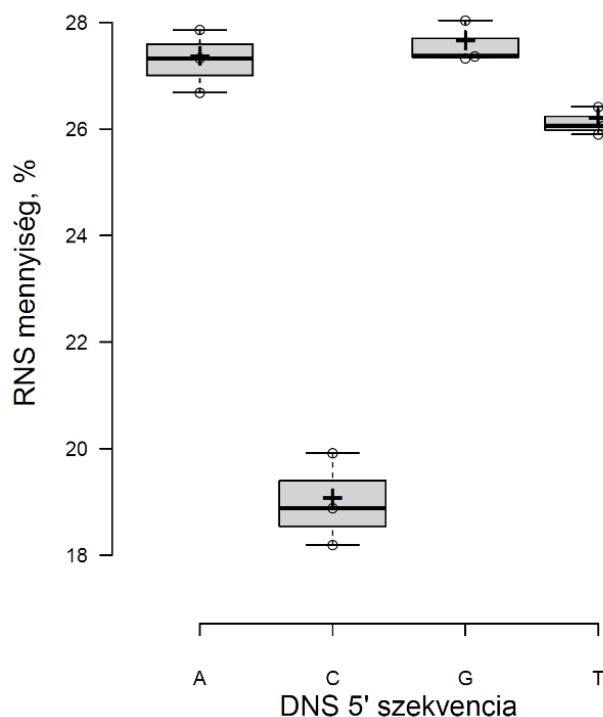


### 52. Ábra. A J23119 promótertől downstream lévő szekvenciák hatása a +1-ről induló transzkripcióra, ha a +1 A, vagy G

Az ábrán a J23119 promóter után elhelyezkedő különböző szekvenciák szerepe látható a +1 pozícióról történő átírára. Az oszlopok megmutatják, hogy átlagban hány %-ban indult el +1-ről az átírás adott templát DNS szekvencia esetén. A hibasávok a százalékos értékek átlagának szórását mutatják. N=3, amely a 3 cDNS könyvtárnak felel meg.



A következő lépésekben azt vizsgáltuk meg, hogy a J23119 promótertől downstream elhelyezkedő nukleotidok hogyan hatnak a transzkriptumok mennyiségére. Az +1-es pozíció szerepét vizsgálva az átíródott RNS-ek mennyiségére azt találtuk, hogy ha abban a bázis purin, több RNS termelődik, mintha az pirimidin lenne (53. ábra). A +1-ben purinokról keletkezett transzkriptumok mennyisége egymástól szignifikánsan nem különbözik. Ennél a promóternél a +1T is megközelíti a purinok szintjét a róla termelődő RNS-ek mennyiségében, és a +1C-ről is körülbelül 2/3-annyi RNS jött létre, mint bármelyik purinról.

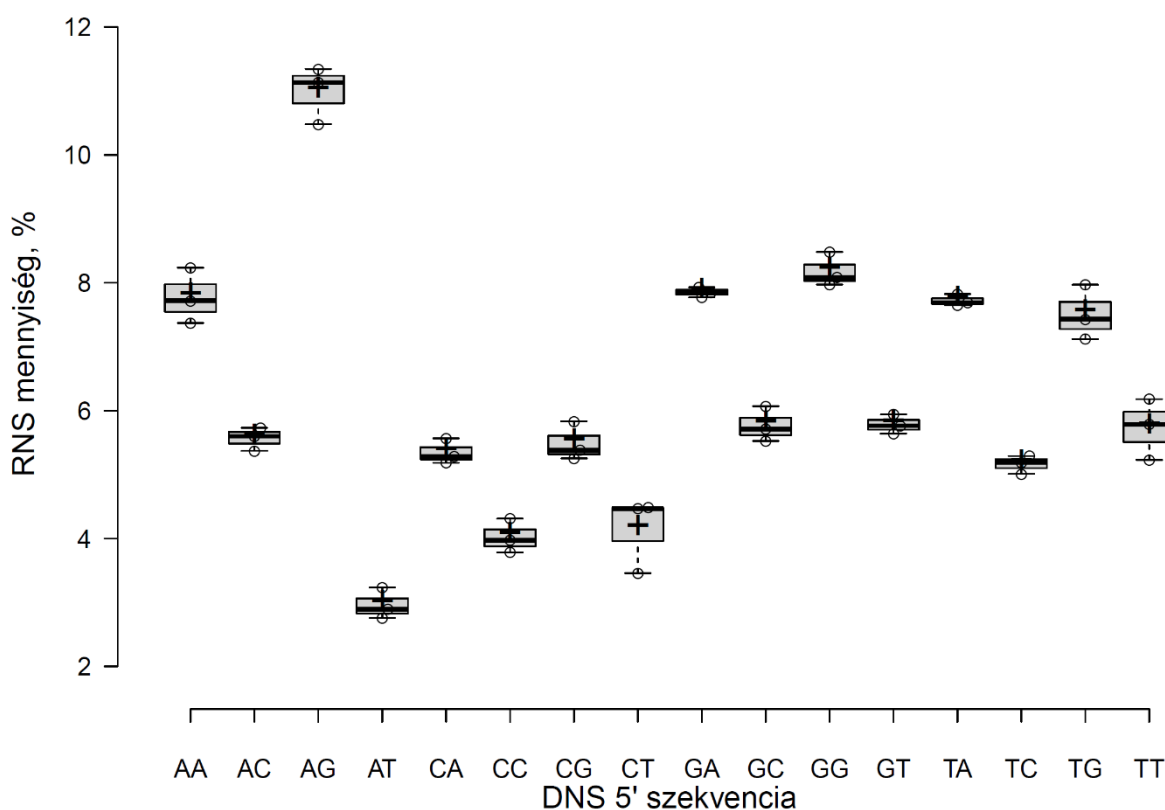


**53. Ábra. A +1-es pozícióban lévő nukleotid hatása az RNS expresszióra**

Az ábrán a +1-es pozícióban lévő nukleotid hatása látható a J23119 promóterről termelődött RNS-ek mennyiségére. Az ábra az összes transzkripció iniciációs pozíció transzkriptumainak mennyiségeit mutatja. 1 adatpont megadja, hogy adott nukleotiddal az összes RNS hány %-a keletkezett az adott cDNS könyvtáron belül. A boxokon belüli vastag vízszintes vonalak a mediánokat mutatják; a boxok határolják az R programmal meghatározott 25. és 75. percentilist; a szimpla vízszintes vonalak az interkvartilis tartomány 1,5-szereséig terjednek ki a 25. és 75. percentilistől; a keresztek a minták átlagát mutatják; az adatpontok körökként vannak ábrázolva. N=3, amely a 3 cDNS könyvtárnak felel meg. A p értékek a Függelékben a 118. oldalon találhatóak meg.

A promótertől downstream lévő első 2 pozíciót vizsgálva azt kaptuk, hogy erről a promóterről keletkezett RNS-ek mennyiségében szintén nagyon nagy szerepet játszik a 2. pozíció is (54. ábra). Szignifikánsan a legtöbb RNS AG kezdetű szekvenciákról jött létre, amiről akár több, mint 3x annyi RNS is keletkezhet, mint a többi +1-ben purint tartalmazó kezdő dinukleotidról. Ezt követően a legtöbb RNS, közel azonos mennyiségekkel akkor keletkezett, ha a promótert AA, GA, GG, TA, vagy TG követte.

Eredményeinkből látszik, hogy +1A-val sokkal nagyobb különbségek vannak a +2-es pozíció függvényében, és +1G és +1T után nagyon hasonló RNS-szintbeli változásokat okoznak a +2-es pozícióban lévő eltérő nukleotidok. Ha a +1G, vagy T, akkor átlagosan kb. 1,5x annyi RNS keletkezett, ha a +2-es pozíció bázisa purin, mint ha pirimidin. Az 54. ábra eredményei alapján keletkezhet olyan mennyiségű RNS a +1-es pozícióban pirimiddal is, mint purinnal, ha megfelelő nukleotid követi, valamint a +1A-val, vagy +1G-vel keletkezett RNS-ek mennyiségében is lehet jelentős eltérés, ha a +2-ben különböző nukleotidok szerepelnek.

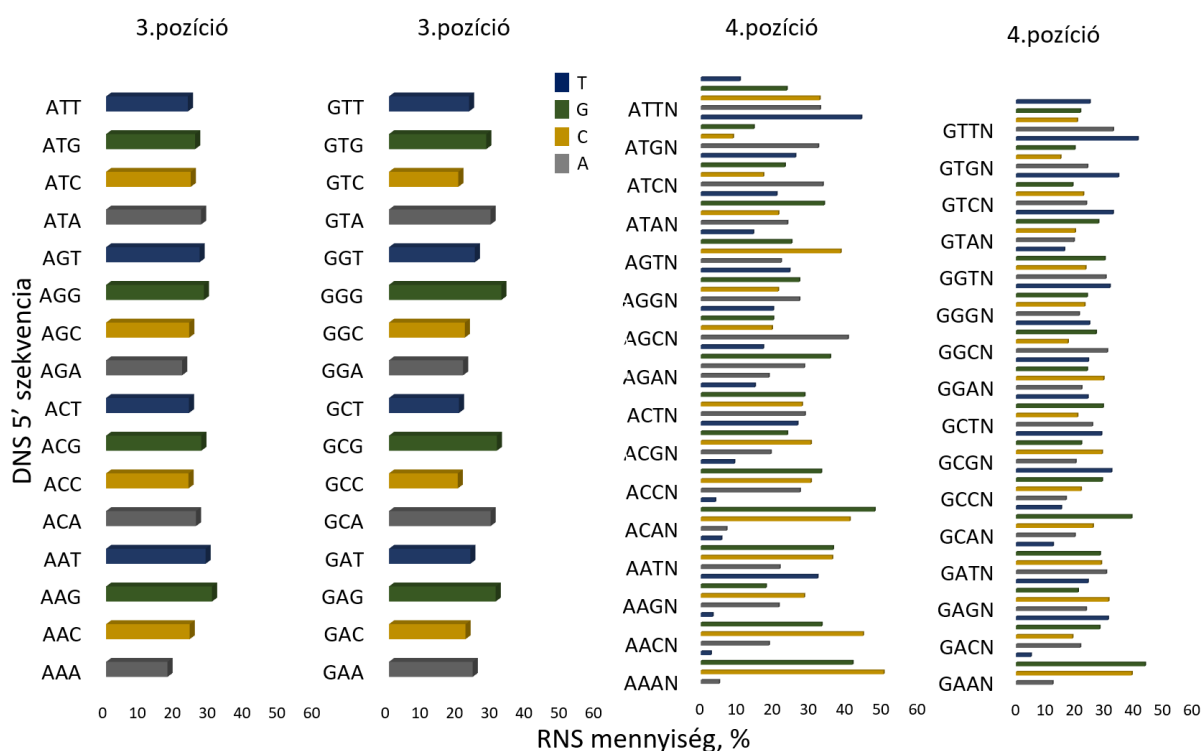


**54. Ábra. A promóterrel downstream elhelyezkedő első két nukleotid hatása az RNS expresszióra**

Az ábrán a promóterrel downstream követő első két nukleotid hatása látható a J23119 promóterrel termelődött RNS-ek mennyiségére. Az ábra az összes transzkripció kezdeti pozíció transzkriptumainak mennyiségét mutatja. 1 adatpont megadja az összes szekvenciáról termelődött RNS-ek összegéhez képest keletkezett RNS-ek százalékos arányát az adott szekvenciáról, adott cDNS könyvtárban belül. A boxokon belüli vastag vízszintes vonalak a mediánokat mutatják; a boxok határolják az R programmal meghatározott 25. és 75. percentilist; a szimpla vízszintes vonalak az interkvartilis tartomány 1,5-szereséig terjednek ki a 25. és 75. percentilistől; a keresztek a minták átlagát mutatják; az adatpontok körökként vannak ábrázolva. N=3, amely a 3 cDNS könyvtárnak felel meg. A p értékek a Függelékben a 118. oldalon találhatóak meg.

Mivel a transzkripció kezdeti pozíciója a purinnal kezdődő szekvenciáknál volt legnagyobb arányban a +1, így a további pozíciók szerepe a disszertáció keretében már csak ezekkel kerül bemutatásra. Az 55. ábrán látható, hogy a +3 és +4 pozíciók is rendkívül fontosak a promóterrel expresszálandó RNS-ek mennyiségének meghatározásában.

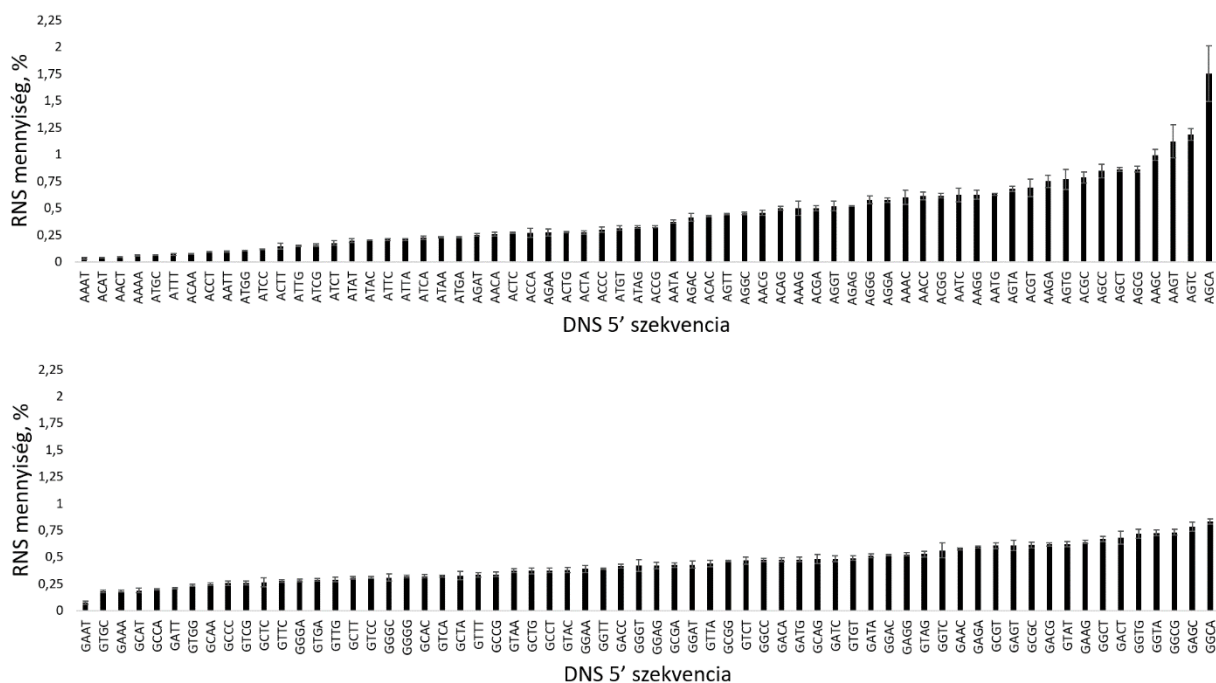
Ha a +1-től +2-ig terjedő pozíciók dinukleotidja megegyezik, viszont az azt követő nukleotid nem, az is szignifikáns különbségeket képes okozni az RNS mennyiségekben, viszont ennek mértéke függ attól, hogy a +1 és +2 pozíciókban milyen szekvencia van. Akár néhányszoros RNS mennyiségbeli különbségeket is képes okozni a +4-es pozícióban elhelyezkedő bázis, ha tőle upstream az első 3 megegyezik, tehát még a +4-es pozícióban lévő nukleotid is befolyásolja a transzkripciót. Az 55. ábrán látható, hogy a +3 és a +4 pozíció nem egyforma módon befolyásolja a transzkriptum mennyiségét, hanem a tőle upstream lévő szekvencia függvényében változóan.



**55. Ábra. A +3-as és +4-es pozícióban lévő nukleotidok hatása az RNS expresszióra**

Az ábrán a +3-as, vagy +4-es pozícióban elhelyezkedő különböző nukleotidok hatása látható a J23119 promóterről termelődött RNS-ek mennyiségére. Az ábra az összes transzkripciós iniciációs pozíció transzkriptumainak mennyiségeit mutatja. 1 oszlop megadja, hogy a promótert downstream követő 3., vagy 4. bázis adott nukleotidjával átlagosan hány % RNS keletkezett az összes, a vizsgált pozíciótól upstream elhelyezkedő pozíciók szekvenciájában megegyező, de a vizsgált pozícióban más nukleotidot tartalmazó szekvenciáról termelődött RNS-ek összegéhez képest. N=3, amely a 3 cDNS könyvtárnak felel meg.

A promótert követő első 4 bázisos szekvencia függvényében a fenti eredményeken látható módon változó mennyiségű RNS expresszálódik a J23119 promóterről. A legkisebb és a legnagyobb expressziót biztosító szekvenciákról termelődött RNS mennyiségek átlagai között több, mint 50x-es különbség van +1A-val (AAAT és AGCA) és közel 5x-ös +1G-vel (GAAT és GGCA) (56. ábra). Eredményeink alapján a J23119 promóternél nemcsak a +1-es pozíció, hanem a továbbiak is befolyásolják a keletkező RNS-ek mennyiségét.



**56. Ábra. A J23119 promótert downstream követő négy bázisos szekvencia hatása az RNS expresszióra, ha a +1 A vagy G**

Az ábrán a +1-től a +4-es pozícióig terjedő szekvencia hatása látható a J23119 promóterről termelődött RNS-ek mennyiségére, ha a +1-es pozícióban A vagy G van. Az ábra az összes transzkripció iniciációs pozíció transzkriptumainak mennyiségeit mutatja. 1 oszlop megadja, hogy az összes lehetséges szekvenciáról (256 variáció) termelődött RNS-ek összegéhez képest adott szekvenciáról az RNS-ek hány %-a keletkezett átlagosan. A hibásávok a százalékos értékek átlagának szórását mutatják. N=3, amely a 3 cDNS könyvtárnak felel meg.

## 6. Diszkusszió

### 6.1. A Cas12a nukleázok

A Cas12a nukleázokat néhány éve publikálták először, mint genomszerkesztésre alkalmas módszert és azért nagyon érdekesek, mert más PAM motívumot ismernek fel, emiatt eltérő targetek hasítására képesek, mint a Cas9 nukleázok, ráadásul eltérő módon, mert az Cas9-el szemben a DNS két szálának hasításakor nem tompa, hanem ragadós végeket hoznak létre. Ezen felül később több csoport is megmutatta, hogy alacsonyabb off-target aktivitással bírnak, mint az SpCas9.

A Cpf1 nukleázokkal kapcsolatosan azt kívántuk megvizsgálni, hogy képesek-e hatékony HR javítás indukálására, illetve milyen a hasítási aktivitásuk a Cas9 nukleázokhoz képest. A nukleázok aktivitásának méréséhez egy GFP esszét alkalmaztam egy korábbi esszé alapján, amellyel a nukleáz hasítása nélkül is kaptunk fluoreszcens GFP jelet. A riporter esszét sikeresen módosítottam úgy, hogy ez a hasítás nélküli GFP háttér csökkent. A létrehozott riporter esszé bármilyen nukleáz aktivitásának vizsgálatára alkalmas lehet, és az elmúlt évek alatt már több száz target és számos nukleáz vizsgálatához alkalmaztunk a csoportunkban, illetve több másik publikációnkban szolgált már a nukleázok aktivitásának becslésére.

A GFxFP esszé azért különösen alkalmas a különböző PAM szekvenciákat felismerő nukleázok aktivitásának becslésére, mert a külső okok kizárásával jobban összemérhetőek egymással. Például egy genomi régiót célozva, a különböző PAM szekvenciák szükségességéből adódóan feltételezhetően más-más genomi környezet, epigenetikai állapot lehetne az egyes targetek között, amely befolyásolhatná a nukleázok aktivitását. Továbbá, a HR javítást nagyban befolyásolja a targetek távolsága a homológ karoktól. Ennek a hatásnak az elkerülése érdekében különböző nukleázok összehasonlításakor a megfelelő PAM-szekvenciákkal rendelkező targeteknek szinte azonos genomi pozícióban kéne elhelyezkedni. Ez azonban messze nem jelent triviális feladatot az eltérő PAM szekvenciák miatt. Az alkalmazott GFP riporter esszé alkalmazásával mindezek kiküszöbölhetőek. Ezen túlmenően a HR javítás hatékonyságának vizsgálata genomi targeteken körülbelül kilenc-tizennégy napig tart, míg a GFxFP esszé elvégzéséhez csak négy nap szükséges, ami további előny. Eredményeink azt mutatják, hogy a két Cas12a nukleáz ugyanolyan, vagy egyes targeteken nagyobb hatékonysággal képes működni, mint az SpCas9 legszélesebb körben alkalmazott Cas9 társai, és aktivitásuk az SpCas9 aktivitásával összemérhető szinten van.

Megmutattuk, hogy a Cas12a nukleázok képesek genomi pozíciókat célozva is hatékony HR javítás kiváltására. Eredményeink alapján a Cas12a nukleázok hatékonyan képesek DNS-rekombinációs javítást indukálni homológ szekvenciák alapján riporter plazmidot célozva és genomi targeteken is az N2a sejtekben- mely más sejtvonal, mint a humán HEK-293FT sejtek, melyben Zetsche és mtsai mutatták be a Cas12a nukleázokkal végzett úttörő kísérleteiket. Ezen kívül ők csak kettő darab targeten, és csak NHEJ javítással vizsgálták a Cas12a nukleázok aktivitását.

Eredményeink arra utalnak, hogy a Cas12a nukleázok fontos helyet foglalnak el az SpCas9 mellett a genomszerkesztési eszközök palettáján. Ezt az elmúlt években számos csoport is megerősítette és megmutatta a Cas12a nukleázok alkalmasságát génszerkesztésre különböző organizmusokban<sup>78</sup>.

## **6.2. A megnövelt specificitású SpCas9 variánsok**

A CRISPR rendszerek egyik legnagyobb hátránya az off-target hatás, melynek kiküszöbölésére irányuló egyik módszer a megnövelt specificitású Cas9 variánsok létrehozása. Kutatásunk egy része, amelyet a disszertációmban is bemutattam, arra irányult, hogy a két, elsőként publikált megnövelt specificitású SpCas9 variáns, az eSpCas9 és a SpCas9-HF1 hasítási aktivitását vizsgáljuk meg a különböző gRNS módosításokkal, melyekre akkor lehet szükség, ha a target nem G-vel kezdődik. Ezek közül a módosítások közül a guideRNS 5' végén elhelyezkedő utolsó nukleotid G-re cserélését, vagy a spacer változatlanul hagyását, továbbá a gRNS-ek 5' végének rövidítését teszteltük le. Azt az eredményt kaptuk, hogy a gRNS-ek 5' végének bármely fent bemutatott módosítása jobban csökkenti a megnövelt specificitású SpCas9 variánsok aktivitását, mint a WT SpCas9-ét. Valószínűleg a megnövelt specificitású Cas9 variánsok azért nem működnek jól az 5'-végükben nem egyező, vagy rövidebb gRNS-ekkel, mert módosításaik következtében kevésbé tolerálják, ha a spacer és a target között eltérés van (azaz érzékenyek a DNS-RNS hibrid strukturális változásaira, amely a fehérjeszerkezet belsejében helyezkedik el). Eredményeink közlése óta már több másik megnövelt specificitású SpCas9 variánst is létrehoztak (lásd. 2.2. fejezet). Eredményeink publikálását követően többen is megerősítették, hogy a megnövelt specificitású SpCas9 variánsok rutinszerűen csak a teljesen egyező 20 nukleotid hosszú spacerekkel használhatók<sup>48,49</sup>. Ahhoz, hogy a megnövelt specificitású nukleáz variánsokat ne csak G-kezdetű targetekkel lehessen hatékonyan használni, több megoldás létezhet. Az egyik, hogy tovább mutálva a nukleázt, jobban elfogadják ezeket a gRNS módosításokat, azonban ez azzal járhat, hogy aktivitásuk a kívánt targeteken is csökken<sup>79</sup>.

A G-vel kezdődő targetek használata azért terjedt el a CRISPR gRNS-ek expressziójára, mert az emlősökben leginkább használt U6 és az *in vitro* használt T7 promóterekről számos csoport azt állította, hogy a promóterektől downstream elhelyezkedő első pozícióban G nukleotid szükséges a hatékony gRNS expresszióhoz. Ezért célszerűnek tűnik megvizsgálni, hogy a gRNS expresszióra általában alkalmazott promóterek esetében megfelelő-e az expresszió más nukleotidokról is, ami által a megnövelt specificitású nukleázokkal alkalmazható targetek köre is bővíthet.

### **6.3. A CRISPR gRNS-ek expressziójára gyakran alkalmazott promóterek**

Mivel a Cas nukleázok működését befolyásolja a gRNS-ek mennyisége és hossza, ezért rendkívül fontos, hogy pontos ismereteink legyenek az alkalmazott promóterekről. Több tanulmány szerint az, hogy milyen nukleotidok vannak a +1-es pozícióban és környezetében, befolyásolhatja az RNS expressziót és a transzkripció iniciáció pozícióját is. Bár néhány általunk vizsgált promóter esetében a +1-es pozícióban elhelyezkedő nukleotidok szerepét már vizsgálták szisztematikusan, de a továbbiakat még nem.

A doktori értekezésem alapjául szolgáló kísérleteinkben a +1-es pozíción túl a promótertől downstream elhelyezkedő további 3 pozícióban található nukleotidok szerepét is megvizsgáltuk az RNS expresszióra és a transzkripció kezdő pozíciójára. Kutatásainkban a CRISPR gRNS-ek expressziójára gyakran alkalmazott promóterek közül a humán U6, a humán 7sk, a T7, az U6T7 és a J23119 promótereket vizsgáltuk meg.

#### **6.3.1. Az U6 promóter**

Az U6 promótert vizsgálva megerősítettük, ha a promótertől downstream elhelyezkedő első pozícióban pirimidin található, csak kis arányban indul a +1-ről a transzkripció. Megmutattuk, hogy ebben az esetben a további pozíciók erősen befolyásolják, hogy melyik pozícióról milyen arányban indul el az átírás, így pirimidinről nagyon heterogén hosszúságú RNS-ek termelődnek. Megerősítettük, ha a +1-es pozícióban purin van, nagyrészt a +1-ről indul el a transzkripció. Megmutattuk, hogy a pirimidinekkel szemben itt a további pozíciók kevésbé befolyásolják a transzkripció iniciáció pozícióját. Megerősítettük, hogy a +1-es pozícióban pirimidinről szignifikánsan kevesebb RNS íródik át, mint purinokról. A +1A-ról és +1G-ről expresszálandó RNS-ek mennyiségében nem találtunk szignifikáns különbséget. Eredményeink alapján az U6 promótertől downstream elhelyezkedő további 3 pozíció szekvenciája is befolyásolja a keletkező RNS-ek mennyiségét.

A továbbiakban abból az okból kifolyólag, hogy +1-es pozícióban pirimidint tartalmazó szekvenciákkal nem nagyrészt a +1-es pozíció volt a transzkripció kezdő helye, már csak a +1-ben purint tartalmazó szekvenciákat vizsgálva azt kaptuk, hogy a +2-es pozíció is fontos szerepet játszik az RNS-ek mennyiségében. A legtöbb RNS az U6 promóterről akkor keletkezik, ha a promótert AG, vagy GG dinukleotid követi, amelyek a legkisebb expressziójú GC kezdetű szekvenciákhoz képest kétszerannyi RNS-t biztosítanak. A promótertől downstream elhelyezkedő 3-4. pozíció a tőle upstream elhelyezkedő szekvencia függvényében változóan befolyásolja az RNS mennyiségét, de akár többszörös RNS mennyiségbeli különbségeket okozhat e pozíciók szekvenciája is. Eredményeink szerint az U6 promótert downstream követő első 4 bázisos szekvenciák függvényében változó mennyiségű RNS expresszálódik az U6 promóterről, és a legkisebb és a legnagyobb expressziót biztosító szekvenciákról termelődött RNS mennyiségek átlagai között egy nagyságrendnyi különbség van.

### **6.3.2. A 7sk promóter**

A 7sk promóterről azt találtuk, hogy promóter utáni szekvenciák hasonlóképp befolyásolják a transzkripció iniciáció pozícióját, mint az U6 promóternél, így míg +1 purinnal a transzkripció iniciációja közel minden esetben a +1, a +1 pirimidinről keletkezett RNS-ek nagy része nem a +1-ről indul, és így többféle hosszúságú RNS keletkezik. Így alaposabb vizsgálatokat a továbbiakban a +1-ben purint tartalmazó szekvenciákkal végeztünk.

Eredményeink alapján +1G-vel összességében mintegy 30%-al több RNS termelődik, mint +1A-val, de az RNS mennyiség függ a további pozíciók szekvenciájától is. Az U6 promóterhez hasonlóan szintén befolyásolja az RNS expressziót a +2- +4-es pozíciók szekvenciája is. A második pozíció tekintetében a legtöbb RNS a 7sk promóterről akkor keletkezik, ha a promótert az AG, a GG és GT dinukleotidok követik, amelyek esetében több, mint kétszer magasabb RNS szint érhető el átlagban, mint a legalacsonyabb RNS szintet biztosító purin kezdetű dinukleotidról. Az RNS mennyiséget a +3-4. pozíciók nukleotidjai itt is az upstream elhelyezkedő szekvenciától függő mértékben befolyásolják. A promótert követő 4 bázisos szekvenciák +1A-val 27x-es, +1G-vel több, mint 7x-es RNS mennyiségbeli különbséget okozhatnak.

Megmutattuk, hogy a 7sk és az U6 promóter hasonló expressziós szintet biztosít HEK-293 sejtekben, így a 7sk akár párhuzamosan, vagy akár az U6 helyett használva is alkalmas lehet gRNS expresszióra.



### 6.3.3. A T7 promóter

A T7 promóter a méréseink szerint az irodalommal egybeesően a +1-es pozícióban a guanint preferálja. Meglepő módon azonban azt találtuk, hogy vannak olyan, a +1-es pozícióban A-t tartalmazó szekvenciák is, amelyek elérik a +1G-ről termelődő RNS-ek mennyiségét és az átírás is nagyrészt +1-ről indul el. A T7 promótert követő 4 nukleotid összes variációjáról keletkezett RNS-eket vizsgálva megállapítottuk, hogy az AG dinukleotidokról kezdődő szekvenciák esetében átlag 84%-ban, míg a +1-ben G-t tartalmazó szekvenciák esetében átlag 90%-ban pontos, +1-ről induló a transzkripció, amit a +2- +4 pozíciók szekvenciája már csak nagyon kis mértékben befolyásol. Érdekes módon azt találtuk, hogy az AG-vel kezdődő szekvenciák átlagosan azonos mennyiségű RNS szintet biztosítanak, mint a +1G-vel legnagyobb RNS szintet biztosító GA kezdetű szekvenciák. Irodalmi adatok alapján nem konzisztens, hogy a T7 promóterhez a promótertől downstream egy, kettő vagy 3 darab G-t érdemes használni. Megmutattuk, hogy a GG és GGG szekvenciákról csak fele- illetve negyedannyi transzkript keletkezik, mint a GA kezdetűekről. Az RNS mennyiséget a +3 és +4 pozíciók nukleotidjai is szignifikánsan, a tőlük upstream elhelyezkedő szekvenciától függő mértékben befolyásolják. A legkisebb és a legnagyobb expressziót biztosító 4 bázisos szekvenciákról termelődött RNS mennyiségek átlagai között közel 5x-ös különbség van AG kezdettel és több, mint 13x-os +1G-vel.

### 6.3.4. Az U6T7 promóter

Az U6T7 hibrid promótert bemutató cikk szerint legalább egy G-vel kell alkalmazni, de a promótertől elhelyezkedő nukleotidok szerepét az RNS expresszióra, vagy a transzkripció iniciációs pozíciókra eddig nem vizsgálták. Az U6T7 promóterrel azt kaptuk, hogy itt is befolyásolják a transzkripció iniciáció pozícióját a promótertől downstream elhelyezkedő nukleotidok. Azonban míg az összes általunk vizsgált promóternél a purinokról, vagy legalább a guaninról majdnem 100%-ban a +1-ről indul el az átírás, az U6T7 promóternél sokkal heterogénebb a transzkripció iniciációjának pozíciója. Az összes promóterhez képest sokkal nagyobb arányban indul el az átírás a -1-ről, még guaninnal is átlagosan 30%-ban. A +1-es pozíciót vizsgálva azt találtuk, hogy még akkor is alacsony arányban indul el +1-ről az átírás, ha az purin, pirimidinről pedig kiemelkedően kis arányban indul el pontosan az átírás a többi promóterrel összehasonlítva. Eredményeink szerint a módosított hibrid promóternél, bár T7 polimerázzal jól használható, emlős expresszióhoz figyelembe kell venni, hogy hiába expresszál az U6, 7sk promóterekhez hasonló mennyiségben RNS-t, de nagyon heterogén hosszúságú RNS-ek keletkeznek róla, még a +1-es pozícióban purinról is.

Ma és munkatársai az U6 promóter olyan módosított változatainál, amelyekben a +1-től upstream elhelyezkedő szekvenciákat módosították hasítóhelyek bevitelének érdekében, jelentős eltéréseket tapasztaltak az RNS iniciáció tekintetében, pontatlanná vált az RNS 5' végének átírása<sup>52</sup>. Esetünkben az U6 és U6T7 utolsó 18 bázisa jelentősen eltér, ezért feltételezhetjük, hogy a fent említett megfigyeléssel megegyező okokból, azaz a +1-es pozíciótól upstream elhelyezkedő szekvenciák módosításának következtében fordulhat elő az, hogy a hibrid promóternek az eredeti promóterhez képest nagymértékben heterogénebb az iniciációs kezdő pozíciója. Feltételezhetően ennek az az oka, hogy ezzel a módosítással flexibilisebbé válik a transzkripció buborék (lásd 6.3.6. fejezet). Eredményeink és gRNS szekvenáló módszerünk felhasználásával érdekes lehetne egy olyan U6T7 promóter megalkotása, amely emlős rendszerekben is homogén hosszúságú RNS-ek transzkripciójára képes.

### **6.3.5. A J23119 promóter**

Tudomásunk szerint a J23119 promóternél eddig nem vizsgálták, hogy milyen hatással vannak a promótertől downstream elhelyezkedő nukleotidok az RNS expresszióra és a transzkripció kezdő helyére. Eredményeink alapján a J23119 promótert downstream követő nukleotidok szintén befolyásolják a kezdő pozíciót. Habár kisebb arányban indul el pontosan az átírás, ha a +1 pirimidin a +1 purinokhoz képest, de a többi vizsgált promóterhez képest több a +1-ből átíródó RNS pirimidinekről is. Ha a +1A vagy G, szinte majdnem mindig a +1-ben kezdődik az átírás. Nincs jelentős eltérés +1A és +1G között abban, hogy milyen pontos a transzkripció iniciációja. A különböző promótertől downstream elhelyezkedő szekvenciák befolyásolják a keletkező RNS-ek mennyiségét is. Megmutattuk, hogy a legtöbb RNS akkor keletkezik, ha a +1-ben purin van. Eredményeink alapján a J23119 promóterről keletkezett RNS mennyiségét is szignifikánsan befolyásolja a +2- +4 pozíciók szekvenciája is. A promótert követő első 4 bázisos szekvencia +1A-val akár 50x-es, +1G-val pedig közel 5x-ös különbséget is okozhat a legjobb és a legrosszabb szekvenciákról termelődött RNS-ek mennyiségében.

### 6.3.6. Konklúzió

Jelenlegi ismereteink szerint egyik általunk vizsgált promóterről sincs a promótertől downstream elhelyezkedő nukleotidok RNS expresszióra és kezdő pozíciókra gyakorolt hatásairól megközelítőleg sem hasonlóan átfogó tanulmány a mi kutatásainkhoz képest. Egy bakteriális promóterről csináltak egy hasonlóan mélyreható vizsgálatot Vedenskaya és mtsai<sup>80</sup>, akik azt kapták, hogy mind a kezdő pozíciókra, mind az RNS expresszióra nézve összességében előnyösebb, ha a +1-ben purin van, de a további pozíciók is számítanak és jellemzően azokban is purin preferenciát tapasztaltak. A doktori értekezésem alapjául szolgáló kísérleteinkben a kiválasztott promóterek sokkal átfogóbb vizsgálatát végeztük el az eddigiekhez képest. Nemcsak a +1-es pozíció, de a tőle downstream elhelyezkedő további 3 bázis hatását is megvizsgáltuk az RNS mennyiségekre és a transzkripció iniciációs pozíciókra. A promóterekre irányuló vizsgálatainkat szekvenciánként nagy lefedettséggel végeztük el a legmodernebb technikák közé tartozó újgenerációs szekvenálás segítségével.

Eredményeink alapján az összes általunk vizsgált promóternél rendkívül nagy szerepet játszik mind az RNS expresszióra, mind a kezdő pozíciókra a promótertől downstream elhelyezkedő szekvencia. Mindkét esetben, mind a 4 promóter utáni pozíció különböző nukleotidjai szignifikáns különbségeket okozhatnak. Kísérleteink alapján a +1-es pozícióban elhelyezkedő nukleotid nem határozza meg a további szekvenciától függetlenül az expressziós szintet, hanem azt a további pozíciók nukleotidjai is jelentősen befolyásolják. Így előfordulhat, hogy nagyságrendbeli különbségek lehetnek az RNS mennyiségekben a +2- +4 pozíciók függvényében, hiába azonos a +1-es pozíció nukleotidja. Mivel az RNS-ek mennyiségét számos külső tényező befolyásolhatja, ezért kísérleteinket az RNS-t kódoló szakaszba épített randomizált bárkódoknak köszönhetően rengeteg különböző RNS szekvenciával végeztük el. Ezzel kívántuk elkerülni a vizsgált 4 bázisos szekvencia expresszióbeli különbségein kívül adódó külső hatásokat, mint például az RNS szerkezetbeli különbségek okozta RNS degradációt.

Több csoport is bemutatta, hogy számos promóternél a promótertől downstream elhelyezkedő nukleotidok a transzkripciós kezdő pozíciókat is befolyásolják<sup>80-85</sup>. A legfrissebb nézet ennek magyarázatára, hogy a transzkripció iniciációjakor a promóternél szétváló kettősszálú DNS, a transzkripciós buborék mérete nem fix. A transzkripciós buborék méretét számos tényező befolyásolhatja, akár a +1-es pozíció környezetében elhelyezkedő szekvencia is, amelyek ahhoz vezethetnek, hogy a transzkripció iniciációja nem egy fix pozícióról történik, hanem különböző arányban akár több pozícióról is, így pedig nem megegyező hosszúságú RNS-ek

keletkezhetnek<sup>80,86</sup>. Az 5 promótert is megvizsgáló kísérleteink alapján megfigyeltük, hogy az összes promóternél befolyásolja a kezdő pozíciókat a promótertől downstream elhelyezkedő szekvencia. Mindegyik promóternél azt kaptuk, hogy ha a promótertől downstream az első pozícióban pirimidin van, sokkal kisebb arányban lesz a +1-es pozíció a kezdő pozíció, és ilyenkor nem egy fix pozícióról indul el az átírás. Abban az esetben, ha a +1-ben pirimidin van, mindegyik promóter esetében nagymértékben befolyásolta nemcsak a promótert követő első, hanem a többi azt követő nukleotid is, hogy melyik kezdő pozícióról milyen arányban indult el az átírás. U6T7 promóteres eredményeink, Ma és mtsai-nak eredményeivel<sup>52</sup> megegyezően azt mutatták, hogy az U6 promóter TATA box utáni módosítása heterogén transzkripció kezdő pozíciókat eredményez, akkor is, ha a +1-es pozícióban purin van. Ebből az következik, hogy a +1-től upstream elhelyezkedő szekvenciák is befolyásolják a transzkripció kezdő pozíciókat.

A promóterekkel kapcsolatos eredményeink több okból is jelentősek lehetnek. Egyrészt, mivel összesen 5 promótert vizsgáltunk meg, melyek 3 különböző organizmusból származnak, ezért azon megállapításaink, hogy a promótertől downstream elhelyezkedő szekvenciák (és nemcsak az első pozíció nukleotidja), mind az RNS expressziót, mind a transzkripció iniciációját befolyásolja, feltételezhetően számos másik promóterre is igazak lehetnek. Ez nemcsak az alkalmazott kutatásokban, hanem az alap kutatás szintjén is nagyon érdekes megállapítás és felveti más promóterek és további pozíciók vizsgálatának szükségességét. Másrészt, eredményeink alapján a vizsgált promótereknél adott szekvenciákkal magasabb expressziós szint, valamint homogénebb hosszúságú transzkriptumok érhetőek el. Ez rendkívül hasznos lehet különböző alkalmazások számára, többek között a CRISPR/Cas rendszerekhez, mert a gRNS-ek 5' végének hossza és a gRNS-ek mennyisége is nagymértékben befolyásolhatja a nukleázok aktivitását. Érdekes lehet akár az RNS interferenciát alkalmazva is, amelyhez szintén gyakran ugyanezeket a promótereket használják. Eddig a CRISPR-es alkalmazásoknál az volt az elterjedt gyakorlat, hogy emlős sejtekben az U6 és 7sk promótereket használva, vagy *in vitro* a T7 promóterrel szinte kizárólag olyan targeteket választanak, amelyekben a spacer 5' első bázisának megfelelő bázis G. Eredményeink alapján a CRISPR-hez választható targetek köre bővíülhet. Eredményeink felhasználásának további érdekes lehetősége- az esetlegesen a további pozíciókra a jövőben lefolytatott eredményekkel együtt- a különböző CRISPR rendszerek hatékonyabb használatához kifejlesztett predikciós programokba való beépítése, melyek segítségével még hatékonyabb célszekvenciák lennének jósolhatók a CRISPR-es alkalmazásokhoz.

## 7. Köszönetnyilvánítás

Köszönetemet elsőként témavezetőmnek, Dr. Welker Ervinnek szeretném kifejezni, amiért lehetőséget nyújtott, hogy a kutatócsoportjában dolgozzak. Nagyon hálás vagyok, hogy az elmúlt években magas színvonalú kutatásokban vehettem részt és számos molekuláris biológiai módszert tanulhattam meg, egy kiváló csapatban.

Köszönettel tartozom a labor jelenlegi és volt munkatársainak: Szűcsné Pulinka Ildikó, Szűcs Judit, Huszár Krisztina, Dr. Tóth Eszter, Dr. Kulcsár Péter István, Tálás András, Krausz Sarah Laura, Ligeti Zoltán, Varga Éva, Dr. Nyeste Antal, Dr. Fodor Elfrieda, Vida István, Bencsura Petra, Dr. Borsy Adrienn, Cserkutiné Czene Bernadett, Pájer-Turgyán Luca Xénia.

Külön köszönet illeti Dr. Tóth Esztert a Cpf1-ekkel kapcsolatos munkájáért és a kooperálásért, Dr. Kulcsár Péter Istvánt és Huszár Krisztinát a disszertációmban bemutatott megnövelt specificitású SpCas9 variánsokkal folytatott munkájukért és Dr. Sturm Ádámot a promóterek vizsgálatával kapcsolatos összes programozási munkájáért.

Hálásan köszönöm az egész Welker-csoportnak, hogy az elmúlt években ilyen remek hangulatú, motiváló légkörben folytathattam doktori kutatásaimat.

Köszönettel tartozom Dr. Szeri Flórának és Dr. Schamberger Anitának, hogy elolvasták és észrevételeikkel jelentősen javították értekezésem színvonalát.

Egész családomnak nagyon szépen köszönöm a rengeteg támogatást, biztatást, türelmet.

## 8. Közreműködések

A 14. ábrán látható eredmények közül a plazmid konstrukciók elkészítésétől a kiértékelésig a 20. pozícióban nem G-t tartalmazó guideRNS-ek megnövelt specificitású SpCas9 variánsokra kifejtett hatásával kapcsolatos vizsgálatok minden lépését teljes mértékben én végeztem el. Az ezzel kapcsolatban a disszertációmban bemutatott további eredményekhez Dr. Kulcsár Péter István és Huszár Krisztina végezték el a többi kísérletet.

A disszertációmban bemutatott Cas12a nukleázok vizsgálatával kapcsolatos kísérletekben Dr. Tóth Eszterrel együtt végeztük el a kísérletek nagy részét, egymás között felosztva a klónozási, sejtenyésztési, mérési és kiértékelési feladatokat, a GFxFP rendszer továbbfejlesztésén és tesztelésén kívül, melyet önállóan végeztem. A sejtfenntartásban, transzfekcióban és a citometriás mérésekben Bencsura Petra is segített. Huszár Krisztina és Dr. Kulcsár Péter István klónozással, Tóth András pedig citometriás méréssel segítette a munkát.

A promóterek vizsgálatával kapcsolatos összes kísérletes munkát, azok kiértékelését és a statisztikát önállóan végeztem. Ez alól kivétel az NGS eredmények kiértékelése, amelyhez Dr. Sturm Ádám nyújtott segítséget, az összes 4.11. fejezetben felsorolt programozási feladatot Pythonban ő végezte el.

## 9. Irodalomjegyzék

- 1 Barrangou, R. & Doudna, J. A. Applications of CRISPR technologies in research and beyond. *Nat Biotechnol* **34**, 933-941, doi:10.1038/nbt.3659 (2016).
- 2 Barman, A., Deb, B. & Chakraborty, S. A glance at genome editing with CRISPR-Cas9 technology. *Curr Genet* **66**, 447-462, doi:10.1007/s00294-019-01040-3 (2020).
- 3 Amitai, G. & Sorek, R. CRISPR-Cas adaptation: insights into the mechanism of action. *Nat Rev Microbiol* **14**, 67-76, doi:10.1038/nrmicro.2015.14 (2016).
- 4 Puschnik, A. S., Majzoub, K., Ooi, Y. S. & Carette, J. E. A CRISPR toolbox to study virus-host interactions. *Nat Rev Microbiol* **15**, 351-364, doi:10.1038/nrmicro.2017.29 (2017).
- 5 Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816-821, doi:10.1126/science.1225829 (2012).
- 6 Komor, A. C., Badran, A. H. & Liu, D. R. CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes. *Cell* **169**, 559, doi:10.1016/j.cell.2017.04.005 (2017).
- 7 Moon, S. B., Kim, D. Y., Ko, J. H. & Kim, Y. S. Recent advances in the CRISPR genome editing tool set. *Exp Mol Med* **51**, 1-11, doi:10.1038/s12276-019-0339-7 (2019).
- 8 Eid, A., Alshareef, S. & Mahfouz, M. M. CRISPR base editors: genome editing without double-stranded breaks. *Biochem J* **475**, 1955-1964, doi:10.1042/BCJ20170793 (2018).
- 9 Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420-424, doi:10.1038/nature17946 (2016).
- 10 Strecker, J. *et al.* RNA-guided DNA insertion with CRISPR-associated transposases. *Science* **365**, 48-53, doi:10.1126/science.aax9181 (2019).
- 11 Anzalone, A. V. *et al.* Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, 149-157, doi:10.1038/s41586-019-1711-4 (2019).
- 12 Hu, J. H. *et al.* Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* **556**, 57-63, doi:10.1038/nature26155 (2018).
- 13 Nishimasu, H. *et al.* Engineered CRISPR-Cas9 nuclease with expanded targeting space. *Science* **361**, 1259-1262, doi:10.1126/science.aas9129 (2018).
- 14 Mout, R., Ray, M., Lee, Y. W., Scaletti, F. & Rotello, V. M. In Vivo Delivery of CRISPR/Cas9 for Therapeutic Gene Editing: Progress and Challenges. *Bioconjug Chem* **28**, 880-884, doi:10.1021/acs.bioconjchem.7b00057 (2017).
- 15 Zuo, Z. & Liu, J. Allosteric regulation of CRISPR-Cas9 for DNA-targeting and cleavage. *Curr Opin Struct Biol* **62**, 166-174, doi:10.1016/j.sbi.2020.01.013 (2020).
- 16 Slaymaker, I. M. *et al.* Rationally engineered Cas9 nucleases with improved specificity. *Science* **351**, 84-88, doi:10.1126/science.aad5227 (2016).
- 17 Jiang, F. *et al.* Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science* **351**, 867-871, doi:10.1126/science.aad8282 (2016).
- 18 Chen, J. S. *et al.* Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. *Nature* **550**, 407-410, doi:10.1038/nature24268 (2017).
- 19 Babu, K. *et al.* Bridge Helix of Cas9 Modulates Target DNA Cleavage and Mismatch Tolerance. *Biochemistry* **58**, 1905-1917, doi:10.1021/acs.biochem.8b01241 (2019).
- 20 Casini, A. *et al.* A highly specific SpCas9 variant is identified by in vivo screening in yeast. *Nat Biotechnol* **36**, 265-271, doi:10.1038/nbt.4066 (2018).
- 21 Lee, J. K. *et al.* Directed evolution of CRISPR-Cas9 to increase its specificity. *Nat Commun* **9**, 3048, doi:10.1038/s41467-018-05477-x (2018).
- 22 Fu, Y., Sander, J. D., Reyon, D., Cascio, V. M. & Joung, J. K. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat Biotechnol* **32**, 279-284, doi:10.1038/nbt.2808 (2014).
- 23 Kim, D., Kim, S., Kim, S., Park, J. & Kim, J. S. Genome-wide target specificities of CRISPR-Cas9 nucleases revealed by multiplex Digenome-seq. *Genome Res* **26**, 406-415, doi:10.1101/gr.199588.115 (2016).
- 24 Yin, H. *et al.* Partial DNA-guided Cas9 enables genome editing with reduced off-target activity. *Nat Chem Biol* **14**, 311-316, doi:10.1038/nchembio.2559 (2018).
- 25 Ryan, D. E. *et al.* Improving CRISPR-Cas specificity with chemical modifications in single-guide RNAs. *Nucleic Acids Res* **46**, 792-803, doi:10.1093/nar/gkx1199 (2018).
- 26 Cromwell, C. R. *et al.* Incorporation of bridged nucleic acids into CRISPR RNAs improves Cas9 endonuclease specificity. *Nat Commun* **9**, 1448, doi:10.1038/s41467-018-03927-0 (2018).
- 27 Guilinger, J. P., Thompson, D. B. & Liu, D. R. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nat Biotechnol* **32**, 577-582, doi:10.1038/nbt.2909 (2014).
- 28 Kim, Y. B. *et al.* Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. *Nat Biotechnol* **35**, 371-376, doi:10.1038/nbt.3803 (2017).
- 29 Li, X. *et al.* Base editing with a Cpf1-cytidine deaminase fusion. *Nat Biotechnol* **36**, 324-327, doi:10.1038/nbt.4102 (2018).
- 30 Kim, D. *et al.* Genome-wide target specificities of CRISPR RNA-guided programmable deaminases. *Nat Biotechnol* **35**, 475-480, doi:10.1038/nbt.3852 (2017).
- 31 Senturk, S. *et al.* Rapid and tunable method to temporally control gene editing based on conditional Cas9 stabilization. *Nat Commun* **8**, 14370, doi:10.1038/ncomms14370 (2017).

32 Zhang, J., Chen, L., Zhang, J. & Wang, Y. Drug Inducible CRISPR/Cas Systems. *Comput Struct Biotechnol J* **17**, 1171-1177, doi:10.1016/j.csbj.2019.07.015 (2019).

33 Shen, B. *et al.* Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. *Nat Methods* **11**, 399-402, doi:10.1038/nmeth.2857 (2014).

34 Kleinstiver, B. P. *et al.* Genome-wide specificities of CRISPR-Cas Cpf1 nucleases in human cells. *Nat Biotechnol* **34**, 869-874, doi:10.1038/nbt.3620 (2016).

35 Toth, E. *et al.* Cpf1 nucleases demonstrate robust activity to induce DNA modification by exploiting homology directed repair pathways in mammalian cells. *Biol Direct* **11**, 46, doi:10.1186/s13062-016-0147-0 (2016).

36 Hur, J. K. *et al.* Targeted mutagenesis in mice by electroporation of Cpf1 ribonucleoproteins. *Nat Biotechnol* **34**, 807-808, doi:10.1038/nbt.3596 (2016).

37 Zetsche, B. *et al.* Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* **163**, 759-771, doi:10.1016/j.cell.2015.09.038 (2015).

38 Vanegas, K. G., Jarczynska, Z. D., Strucko, T. & Mortensen, U. H. Cpf1 enables fast and efficient genome editing in *Aspergilli*. *Fungal Biol Biotechnol* **6**, 6, doi:10.1186/s40694-019-0069-6 (2019).

39 Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* **31**, 827-832, doi:10.1038/nbt.2647 (2013).

40 Ryan, O. W. *et al.* Selection of chromosomal DNA libraries using a multiplex CRISPR system. *Elife* **3**, doi:10.7554/eLife.03703 (2014).

41 Min, Y. L. *et al.* CRISPR-Cas9 corrects Duchenne muscular dystrophy exon 44 deletion mutations in mice and human cells. *Sci Adv* **5**, eaav4324, doi:10.1126/sciadv.aav4324 (2019).

42 Schoger, E. *et al.* CRISPR-Mediated Activation of Endogenous Gene Expression in the Postnatal Heart. *Circ Res* **126**, 6-24, doi:10.1161/CIRCRESAHA.118.314522 (2020).

43 Dahlman, J. E. *et al.* Orthogonal gene knockout and activation with a catalytically active Cas9 nuclease. *Nat Biotechnol* **33**, 1159-1161, doi:10.1038/nbt.3390 (2015).

44 Kiani, S. *et al.* Cas9 gRNA engineering for genome editing, activation and repression. *Nat Methods* **12**, 1051-1054, doi:10.1038/nmeth.3580 (2015).

45 Tsai, S. Q. *et al.* GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat Biotechnol* **33**, 187-197, doi:10.1038/nbt.3117 (2015).

46 Wienert, B. *et al.* Unbiased detection of CRISPR off-targets in vivo using DISCOVER-Seq. *Science* **364**, 286-289, doi:10.1126/science.aav9023 (2019).

47 Cho, S. W. *et al.* Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res* **24**, 132-141, doi:10.1101/gr.162339.113 (2014).

48 Kulcsar, P. I. *et al.* Crossing enhanced and high fidelity SpCas9 nucleases to optimize specificity and cleavage. *Genome Biol* **18**, 190, doi:10.1186/s13059-017-1318-8 (2017).

49 Zhang, D. *et al.* Perfectly matched 20-nucleotide guide RNA sequences enable robust genome editing using high-fidelity SpCas9 nucleases. *Genome Biol* **18**, 191, doi:10.1186/s13059-017-1325-9 (2017).

50 Kim, S., Bae, T., Hwang, J. & Kim, J. S. Rescue of high-specificity Cas9 variants using sgRNAs with matched 5' nucleotides. *Genome Biol* **18**, 218, doi:10.1186/s13059-017-1355-3 (2017).

51 Gao, Z., Harwig, A., Berkhout, B. & Herrera-Carrillo, E. Mutation of nucleotides around the +1 position of type 3 polymerase III promoters: The effect on transcriptional activity and start site usage. *Transcription* **8**, 275-287, doi:10.1080/21541264.2017.1322170 (2017).

52 Ma, H. *et al.* Pol III Promoters to Express Small RNAs: Delineation of Transcription Initiation. *Mol Ther Nucleic Acids* **3**, e161, doi:10.1038/mtna.2014.12 (2014).

53 Helm, M., Brule, H., Giege, R. & Florentz, C. More mistakes by T7 RNA polymerase at the 5' ends of in vitro-transcribed RNAs. *RNA* **5**, 618-621, doi:10.1017/s1355838299982328 (1999).

54 Herrera-Carrillo, E., Gao, Z. L., Harwig, A., Heemskerk, M. T. & Berkhout, B. The influence of the 5-terminal nucleotide on Ago-sRNA activity and biogenesis: importance of the polymerase III transcription initiation site. *Nucleic Acids Res* **45**, 4036-4050, doi:10.1093/nar/gkw1203 (2017).

55 Romanienko, P. J. *et al.* A Vector with a Single Promoter for In Vitro Transcription and Mammalian Cell Expression of CRISPR gRNAs. *PLoS One* **11**, e0148362, doi:10.1371/journal.pone.0148362 (2016).

56 Imburgio, D., Rong, M., Ma, K. & McAllister, W. T. Studies of promoter recognition and start site selection by T7 RNA polymerase using a comprehensive collection of promoter variants. *Biochemistry* **39**, 10419-10430, doi:10.1021/bi000365w (2000).

57 Hwang, W. Y. *et al.* Heritable and precise zebrafish genome editing using a CRISPR-Cas system. *PLoS One* **8**, e68708, doi:10.1371/journal.pone.0068708 (2013).

58 Peeters, B. & de Leeuw, O. A single-plasmid reverse genetics system for the rescue of non-segmented negative-strand RNA viruses from cloned full-length cDNA. *J Virol Methods* **248**, 187-190, doi:10.1016/j.jviromet.2017.07.008 (2017).

59 Kruspe, S. & Giangrande, P. H. Design and Preparation of Aptamer-siRNA Chimeras (AsiCs) for Targeted Cancer Therapy. *Methods Mol Biol* **1632**, 175-186, doi:10.1007/978-1-4939-7138-1\_11 (2017).

60 Kopsidas, G. *et al.* RNA mutagenesis yields highly diverse mRNA libraries for in vitro protein evolution. *BMC Biotechnol* **7**, 18, doi:10.1186/1472-6750-7-18 (2007).

61 Goldrick, M. M. RNase cleavage-based methods for mutation/SNP detection, past and present. *Hum Mutat* **18**, 190-204, doi:10.1002/humu.1175 (2001).

62 Wang, R., Cotten, S. W. & Liu, R. mRNA display using covalent coupling of mRNA to translated proteins. *Methods Mol Biol* **805**, 87-100, doi:10.1007/978-1-61779-379-0\_6 (2012).



63 Luo, M. L., Mullis, A. S., Leenay, R. T. & Beisel, C. L. Repurposing endogenous type I CRISPR-Cas systems for  
programmable gene repression. *Nucleic Acids Res* **43**, 674-681, doi:10.1093/nar/gku971 (2015).

64 Ding, T. *et al.* Reversed paired-gRNA plasmid cloning strategy for efficient genome editing in *Escherichia coli*.  
*Microb Cell Fact* **19**, 63, doi:10.1186/s12934-020-01321-4 (2020).

65 Hawkins, J. S., Wong, S., Peters, J. M., Almeida, R. & Qi, L. S. Targeted Transcriptional Repression in Bacteria  
Using CRISPR Interference (CRISPRi). *Methods Mol Biol* **1311**, 349-362, doi:10.1007/978-1-4939-2687-9\_23  
(2015).

66 Spencer, J. M. & Zhang, X. Deep mutational scanning of *S. pyogenes* Cas9 reveals important functional domains.  
*Sci Rep* **7**, 16836, doi:10.1038/s41598-017-17081-y (2017).

67 Liow, L. T., Go, M. D. K. & Yew, W. S. Characterisation of Constitutive Promoters from the Anderson library in  
*Chromobacterium violaceum* ATCC 12472. *Engineering Biology* **3**, 57-66, doi:10.1049/enb.2018.5007 (2019).

68 Nelson, C. E. *et al.* In vivo genome editing improves muscle function in a mouse model of Duchenne muscular  
dystrophy. *Science* **351**, 403-407, doi:10.1126/science.aad5143 (2016).

69 Esvelt, K. M. *et al.* Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nat Methods* **10**, 1116-  
1121, doi:10.1038/nmeth.2681 (2013).

70 Mashiko, D. *et al.* Generation of mutant mice by pronuclear injection of circular plasmid expressing Cas9 and  
single guided RNA. *Sci Rep* **3**, 3355, doi:10.1038/srep03355 (2013).

71 Jacobus, A. P. & Gross, J. Optimal cloning of PCR fragments by homologous recombination in *Escherichia coli*.  
*PLoS One* **10**, e0119221, doi:10.1371/journal.pone.0119221 (2015).

72 Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* **6**, 343-  
345, doi:10.1038/nmeth.1318 (2009).

73 Doench, J. G. *et al.* Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-  
Cas9. *Nat Biotechnol* **34**, 184-191, doi:10.1038/nbt.3437 (2016).

74 Chari, R., Mali, P., Moosburner, M. & Church, G. M. Unraveling CRISPR-Cas9 genome engineering parameters  
via a library-on-library approach. *Nat Methods* **12**, 823-826, doi:10.1038/nmeth.3473 (2015).

75 Goomer, R. S. & Kunkel, G. R. The transcriptional start site for a human U6 small nuclear RNA gene is dictated by  
a compound promoter element consisting of the PSE and the TATA box. *Nucleic Acids Res* **20**, 4903-4912,  
doi:10.1093/nar/20.18.4903 (1992).

76 Kuhn, H. & Frank-Kamenetskii, M. D. Template-independent ligation of single-stranded DNA by T4 DNA ligase.  
*FEBS J* **272**, 5991-6000, doi:10.1111/j.1742-4658.2005.04954.x (2005).

77 Kwok, C. K., Ding, Y., Sherlock, M. E., Assmann, S. M. & Bevilacqua, P. C. A hybridization-based approach for  
quantitative and low-bias single-stranded DNA ligation. *Anal Biochem* **435**, 181-186, doi:10.1016/j.ab.2013.01.008  
(2013).

78 Paul, B. & Montoya, G. CRISPR-Cas12a: Functional overview and applications. *Biomed J* **43**, 8-17,  
doi:10.1016/j.bj.2019.10.005 (2020).

79 Kulcsar, P. I. *et al.* Blackjack mutations improve the on-target activities of increased fidelity variants of SpCas9  
with 5'-G-extended sgRNAs. *Nat Commun* **11**, 1223, doi:10.1038/s41467-020-15021-5 (2020).

80 Vvedenskaya, I. O. *et al.* Massively Systematic Transcript End Readout, "MASTER": Transcription Start Site  
Selection, Transcriptional Slippage, and Transcript Yields. *Mol Cell* **60**, 953-965,  
doi:10.1016/j.molcel.2015.10.029 (2015).

81 Nepal, C. *et al.* Dynamic regulation of the transcription initiation landscape at single nucleotide resolution during  
vertebrate embryogenesis. *Genome Res* **23**, 1938-1950, doi:10.1101/gr.153692.112 (2013).

82 Consortium, F. *et al.* A promoter-level mammalian expression atlas. *Nature* **507**, 462-470,  
doi:10.1038/nature13182 (2014).

83 Chen, R. A. *et al.* The landscape of RNA polymerase II transcription initiation in *C. elegans* reveals promoter and  
enhancer architectures. *Genome Res* **23**, 1339-1347, doi:10.1101/gr.153668.112 (2013).

84 Yamashita, R. *et al.* Genome-wide characterization of transcriptional start sites in humans by integrative  
transcriptome analysis. *Genome Res* **21**, 775-789, doi:10.1101/gr.110254.110 (2011).

85 Hoskins, R. A. *et al.* Genome-wide analysis of promoter architecture in *Drosophila melanogaster*. *Genome Res* **21**,  
182-192, doi:10.1101/gr.112466.110 (2011).

86 Robb, N. C. *et al.* The transcription bubble of the RNA polymerase-promoter open complex exhibits  
conformational heterogeneity and millisecond-scale dynamics: implications for transcription start-site selection. *J*  
*Mol Biol* **425**, 875-885, doi:10.1016/j.jmb.2012.12.015 (2013).

## 10. Összefoglalás

A *Streptococcus pyogenes* baktériumból származó Cas9 endonukleáz (SpCas9) a génmódosításra jelenleg legelterjedtebben alkalmazott CRISPR/Cas rendszer. A specificitást a guideRNS (gRNS) 5' végén elhelyezkedő 20 bázispáros szekvencia, a spacer biztosítja. A spacer segítségével kerül felismerésre a célszekvencia komplementer DNS-RNS hibridizáció alapján. A nukleáz a DNS mindkét szálát hasítja. Ennek hatására DNS javítási útvonalak indulnak be, melyek a DNS specifikus módosítására használhatók fel. A 2 fő útvonal a homológ rekombinációs (HR) és a nem homológ végcsatlakozás (NHEJ) hibajavítási mechanizmusok. A célszekvencia felismeréséhez és hasításához szükség van egy néhány bázispáros szekvencia jelenlétére is a target mellett, ez a PAM szekvencia. A mai molekuláris biológiai technikákkal már könnyedén változtatható a spacer szekvenciája, ennek köszönhető a rendszer robbanásszerű elterjedése. A Cas9 fehérje segítségével szinte bármilyen genomi szakasz szabadon módosítható, ezért jelentős biotechnológiai, orvosi és tudományos felhasználása van. A Cas12a nukleázokat néhány éve publikálták először, mint genomszerkesztésre alkalmas módszert. Amiatt nagyon érdekesek, mert más PAM szekvenciát ismernek fel, mint az SpCas9, így eltérő targetek hasítására képesek, ráadásul eltérő módon, mert az Cas9-el szemben a DNS két szálának hasításakor nem tompa, hanem ragadós végeket hoznak létre. Azt is megmutatták, hogy alacsonyabb off-target aktivitással bírnak, mint az SpCas9. A CRISPR/Cas rendszereknek vannak limitációi. Az egyik ilyen az off-target hatás, azaz a célszekvenciákhoz hasonló, nem kívánt targetek hasítása. Ennek megoldására számos törekvés van, az egyik ezek közül a nukleáz mutációkkal történő módosítása specifikus helyeken. Napjainkig már több olyan SpCas9 variánst hoztak létre, amelyek megnövekedett specificitással bírnak.

A gRNS-ek előállítására két elterjedt, az RNS szintetizálathatásnál költséghatékonyabb módszer van a gyakorlatban. Az egyik a T7 promóter segítségével történő *in vitro* RNS transzkripció, a másik a különböző organizmusokban más-más kis RNS expresszióra alkalmas promóterek használata. Számos kutatás szerint a Cas nukleázok működését befolyásolja a gRNS-ek mennyisége és hossza, ezért rendkívül fontos, hogy pontos ismereteink legyenek az alkalmazott promóterekről. Több tanulmány szerint a promótertől downstream elhelyezkedő nukleotidok befolyásolhatják az RNS expressziót és a transzkripció iniciáció pozícióját is. Bár néhány általunk vizsgált promóter esetében a promótert követő első bázis szerepét már vizsgálták szisztematikusan, de a további nukleotidok szerepét még nem.

Ezen vizsgálatok alapján a humán 7sk és U6 promóterek egyaránt alkalmazhatók a +1-es pozícióban A-val és G-vel és a várt hosszúságú RNS-eket kapták róluk.

A T7 promóterről ismert, hogy a promótertől downstream elhelyezkedő első bázis befolyásolja a transzkripció iniciációjának pozícióját, és az eddigi eredmények szerint legalább egy G-t kell használni a megfelelő expresszióhoz. Az azonban az irodalom szerint nem konzisztens, hogy a T7 promóterhez egy, kettő vagy 3 darab guanint szükséges alkalmazni.

Az U6T7 hibrid promótert az U6 promóter módosításával hozták létre, úgy, hogy a promóter végét a T7 promóterre cserélték le. A hibrid promótert bemutató cikk szerint a +1-ben legalább egy G-vel ajánlatos használni.

A bakteriális J23119 promóterekkel kapcsolatban tudomásunk szerint eddig nem került publikálásra a +1-es pozícióval kapcsolatos eredmény.

Habár az U6 és 7sk promóterekről megmutatták, hogy a +1-es pozícióban A-val is hatékony expresszióra képesek, a gyakorlatban azonban az SpCas9 gRNS-ekhez általában olyan targeteket választanak, amelyekben a spacer 5' első bázisának megfelelő bázis G, amely a targetek számát szűkíti. Megoldásképp különböző gRNS módosításokat alkalmaznak, amelyek viszont befolyásolhatják a nukleázok aktivitását.

Ezek a +1-es pozícióra irányuló vizsgálatok általában nem a jelenleg a kis RNS-ek hosszáról és mennyiségéről legpontosabb képet adó újgenerációs szekvenálással (NGS) történtek, hanem egyéb kevésbé pontos módszerekkel, illetve az eddig napvilágot látott NGS-el kapott eredményekhez nagyon kis minta számot alkalmaztak.

Doktori értekezésemben a Cas12a nukleázokkal kapcsolatosan arra kerestük a választ, hogy képesek-e hatékony HR javítás indukálására N2a sejtekben az As- és LbCas12a nukleázok, és mennyire hatékonyak a leggyakrabban alkalmazott Cas9 nukleázokhoz képest.

Összehasonlítottuk az As-, és LbCas12a nukleázok hatékonyságát az Sp- és 3 másik, gyakran alkalmazott Cas9 nukleáz aktivitásával. A nukleázok hasítási aktivitásának vizsgálatához a GFxFP nevű esszét optimalizáltuk egy korábban publikált esszéből, ahol a nukleáz hasítása nélkül is kaptunk háttér fluoreszcens GFP jelet. A riporter esszét sikeresen módosítottam úgy, hogy ez a hasítás nélküli GFP háttér csökkent. Az esszé használatával a különböző PAM szekvenciákból adódó eltérő genomi pozíciók okozta aktivitásbeli különbségeket kívántuk kizárni. Eredményeink alapján bár az SpCas9 hatékonyabbnak bizonyult: nagyobb arányban váltott ki HR javítást a sejtekben, mint a Cas12a-k, de a többi Cas9-el viszont hasonló mértékben, illetve egyes targeteken jobban teljesítettek a Cas12a nukleázok. Nemcsak riporter esszével, de genomi célpontokon is teszteltük a Cas12a nukleázok alkalmasságát

genomszerkesztésre. Eredményeink szerint alkalmasak HR javítás indukálására genomi targeteket célozva is. A Cas12a nukleázokkal kapcsolatos vizsgálataink során nem a Zetsche és mtsai. által használt emlős sejtekben, hanem más emlős sejteken (N2a) mutattuk be, hogy a Cas12a nukleázok alkalmasak genomszerkesztésre, ráadásul míg ők csak az NHEJ javítási útvonal segítségével tesztelték le aktivitásukat, mi HR javítást indukáltunk a genomban, valamint a GFxFP plazmidot célozva. Eredményeink alapján a Cas12a nukleázok hasznos alternatívát nyújthatnak a genomszerkesztő eszközök palettáján, ami egybeesik azzal, hogy az elmúlt években számos csoport szintén megmutatta alkalmasságukat a génszerkesztésre többféle organizmusban is.

A megnövelt specificitású eSpCas9 és SpCas9HF1 variánsokkal kapcsolatban azt kívántuk megtudni, hogy aktivitásukat hogyan befolyásolják a különböző gRNS módosítások, amelyeket akkor alkalmaznak, ha a targetben a gRNS első átíródo bázisának megfelelő bázis nem guanin. Ezek közül a módosítások közül a gRNS 5' végén elhelyezkedő utolsó nukleotid G-re cserélését, vagy a spacer változatlanul hagyását, továbbá a gRNS-ek 5' végének rövidítését teszteltük le. Hatásukat egy EGFP-t tartalmazó egér neuroblasztóma sejtvonalban, az EGFP hasítás hatékonyságával teszteltük. Eredményeink alapján, míg a vad típusú SpCas9 aktivitását csak kis mértékben befolyásolják ezek a módosítások, addig a nagy specificitású Cas9 variánsok aktivitását különböző mértékben, de jobban rontják. Így eredményeink alapján e nukleáz variánsokat ilyen módosításokkal nem lehet rutinszerűen alkalmazni. Ennek feloldására többféle megoldás is lehetséges, az egyik a nukleáz variánsok olyan további mutációkkal történő módosítása, amely által jobban tolerálják a gRNS módosításait. A megnövelt specificitású nukleázok targeteinek körét bővítendő felmerül az igény a kis RNS expresszióra alkalmas promóterek mélyrehatóbb vizsgálatára. Ennek segítségével lehetne olyan egyéb szekvenciákat találni, amelyekkel gRNS módosítások nélkül is használhatóak lennének ezek a nukleázok.

A disszertációmban bemutatott munka nagyobbik részében arra kerestünk választ, hogy a gRNS-ekhez gyakran alkalmazott promótereknél a promótertől downstream elhelyezkedő bázisok milyen módon befolyásolják az RNS expressziót és a kezdő szekvenciákat.

A doktori értekezésem alapjául szolgáló kísérleteinkben a kiválasztott promóterek sokkal átfogóbb vizsgálatát végeztük el az eddigiekhez képest. Nemcsak a +1-es pozíció, de a tőle downstream elhelyezkedő további 3 pozíció szerepét is megvizsgáltuk az RNS mennyiségekre és a transzkripció iniciációs pozíciókra. Kutatásunkban minden promóterhez plazmid

könyvtárakat hoztunk létre, és a plazmid könyvtárakról expresszáldott RNS-ek szekvenálásához egy egyedi, gRNS-ek szekvenálására alkalmas módszert alkalmaztunk. Kísérleteinket az RNS-t kódoló szakaszba épített randomizált bárkódoknak köszönhetően rengeteg különböző RNS szekvenciával végeztük el, amellyel a vizsgált 4 bázisos szekvencia különbségein kívül adódó külső hatásokat, mint például szekvenciaspecifikus RNS degradáció, kívántuk elkerülni. A promóterekre irányuló vizsgálatainkat 4 bázispáros szekvenciánként nagy lefedettséggel végeztük el a legmodernebb technikák közé tartozó újgenerációs szekvenálás segítségével.

Az NGS eredményeink alapján az összes vizsgált promóterre igaz, hogy a promótertől downstream lévő pozíciók nukleotidjai befolyásolhatják a keletkező RNS-ek mennyiségét és a transzkripció iniciációjának pontos helyét, ezáltal az RNS hosszát is. Eddig csak a +1-es pozíciót vizsgálták szisztematikusan, azt is csak a T7, U6 és 7sk promóterek esetében. Eredményeink alapján nemcsak a +1 pozíció, de a tőle downstream lévő pozíciók is fontosak az RNS expresszió szempontjából. Így előfordulhat, hogy nagyságrendbeli különbségek lehetnek az RNS mennyiségekben a további pozíciók függvényében, hiába azonos a +1-es pozíció nukleotidja.

Azt találtuk, hogy az összes vizsgált promóternél befolyásolhatja a transzkripció kezdő pozícióját is a promótertől downstream elhelyezkedő szekvencia. Mindegyik promóternél azt kaptuk, hogy ha a promótertől downstream az első pozícióban pirimidin van, sokkal kisebb arányban lesz a +1-es pozíció a kezdő pozíció, és ilyenkor nem egy fix pozícióról indul el az átírás. Abban az esetben, ha a +1-ben pirimidin van, mindegyik promóter esetében nagymértékben befolyásolta nemcsak a promótert követő első, hanem a többi azt követő nukleotid is, hogy melyik kezdő pozícióról milyen arányban indult el az átírás.

Megmutattuk, hogy a humán U6 promóterről expresszáldó RNS-ek mennyiségében a +1A és a +1G között nincs szignifikáns különbség, ami egybeesik azokkal a publikációkkal, amelyek szerint az U6 promóter +1A-val is használható. Ha a +1 pozícióban A, vagy G nukleotid van, a transzkripció kezdő pozíciója közel minden esetben a +1, melyet jelentős módon már nem befolyásol a +2- +4 pozíciók szekvenciája.

Azt találtuk, hogy a humán 7sk promóter szintén nemcsak +1 guaninnal, hanem adeninnal is használható kis RNS expresszióra. A 7sk promóternél a +1-es pozícióban a purin jelenléte pontos, +1 pozíciójú transzkripciós iniciációt okoz, amin a további +2- +4. pozíciók szekvenciája jelentősen nem változtat. +1G-vel szignifikánsan több RNS keletkezik, mint +1A-val. Azt találtuk, hogy HEK sejtekben az U6 és a 7sk promóterekről termelődött RNS-ek

mennyiségében nincs szignifikáns különbség, így a 7sk párhuzamosan, vagy akár az U6 promóter helyett is alkalmas lehet gRNS expresszióra.

A T7 promóter a méréseink szerint az eddigi adatokkal megegyezően a +1-es pozícióban a guanint preferálja. Vannak azonban olyan adeninnel kezdődő szekvenciák is, amelyek megközelítik a +1G-ről termelődő RNS-ek mennyiségét és az átírás is nagyrészt +1-ről indul el. Eredményeink szerint a +1-ben G-t tartalmazó szekvenciák közül a GG és GGG kezdetűekről kevesebb transzkript keletkezik, mint a nagyobb expressziót biztosító GA kezdetűekről.

Eredményeink alapján a módosított hibrid U6T7 promóter, bár T7 polimerázzal jól használható, emlős expresszióhoz figyelembe kell venni, hogy bár azt találtuk, hogy az U6T7, a 7sk és az U6 promóterekről keletkezett RNS-ek mennyiségében nincs szignifikáns különbség, de nagyon heterogén hosszúságú RNS-ek keletkeznek róla, még +1purinról is.

Megvizsgáltuk a J23119 promótertől downstream elhelyezkedő szekvencia hatását a transzkripcióra, és azt találtuk, hogy ha a +1-es pozícióban purin van, a transzkripció iniciáció helye majdnem mindig a +1-es pozíció, amit a +2- +4 pozíciók szekvenciája már nem befolyásol jelentős mértékben. Továbbá megmutattuk, hogy a legtöbb RNS akkor keletkezik, ha a +1-ben purin van.

A promóterekkel kapcsolatos eredményeink több okból is jelentősek lehetnek.

A 3 különböző organizmusból származó promóterek mindegyikénél azt találtuk, hogy a promótertől downstream nemcsak az első, hanem további 3 pozíció is befolyásolja a transzkripció iniciációs pozíciót, ezáltal az RNS-ek hosszát, valamint az expressziót is. Ez az alap kutatásban is érdekes lehet és felveti más promóterek, illetve további pozíciók vizsgálatának szükségességét.

Eredményeink alapján a vizsgált promóterekkel adott szekvenciákat alkalmazva nagyobb RNS szint és pontosabb hosszúságú RNS-ek érhetőek el. Ez rendkívül hasznos lehet különböző alkalmazások számára, mint például az RNS interferencia, vagy a CRISPR. Eredményeink alapján a CRISPR-hez választható targetek köre bővíthet, abból adódóan, hogy találtunk olyan a promóterektől downstream elhelyezkedő szekvenciákat, amelyekről eddig nem volt ismert, hogy ugyanolyan alkalmasak lehetnek gRNS expresszióra, mint az eddig ismert variációk. Eredményeink ezen kívül- akár további pozíciók vizsgálatából származó eredményekkel együtt- alkalmasak lehetnének a jelenlegi CRISPR target predikciós programok továbbfejlesztéséhez is.

A disszertációmban bemutatott eredmények hozzájárulhatnak a CRISPR/Cas rendszer hatékonyabb alkalmazásához.

## 11. Summary

The Cas9 endonuclease (SpCas9) from the bacteria *Streptococcus pyogenes* is currently the most widely applied CRISPR/Cas system used for genetic engineering. Its specificity is provided by the 20 bp ‘spacer’ sequence at the 5' end of the guideRNA (gRNA). The spacer is used to identify the target sequence based on complementary DNA-RNA hybridization. Nuclease cleaves both strands of DNA. This triggers DNA repair pathways that can be used to specifically modify DNA. The 2 main pathways are homologous recombination (HR) and non-homologous end joining (NHEJ) error correction mechanisms. Recognition and cleavage of the target sequence also requires the presence of a few base pairs in addition to the target, the PAM sequence. With today's molecular biology techniques, the sequence of the spacer can be easily changed, thanks to which the system has explosively spread. Almost any section of the genome can be freely modified with the Cas9 protein, so it has significant biotechnological, medical and scientific uses.

Cas12a nucleases were first published a few years ago as a method for genome editing. They are very interesting because they recognize a different PAM sequence than SpCas9, so they are able to cleave different targets, and in a different way, because in contrast to Cas9, cleavage of two strands of DNA creates sticky ends rather than blunt ones. They have also been shown to have lower off-target activity than SpCas9. CRISPR/Cas systems have limitations. One of these is the off-target effect, i.e. the cleavage of unwanted targets similar to the target sequences. There were several attempts to address this, one of which is to modify the nuclease with mutations at specific sites. To date, several SpCas9 variants have been developed that have increased specificity.

There are two common methods for producing gRNAs in practice than the more cost-effective RNA synthesis. One is using the T7 promoter *in vitro* for RNA transcription, and the other is the use of promoters suitable small RNA expression in different organisms. Numerous studies have shown that the activity of Cas nucleases is influenced by the amount and length of gRNAs, so it is extremely important to have an accurate knowledge of the promoters used.

Several studies suggest that nucleotides downstream of the promoter can also affect RNA expression and the position of transcription initiation. Although for some of the promoters we studied, the role of the first nucleotide downstream of the promoter has already been systematically investigated, but the role of additional nucleotides has not yet been studied.

Based on these studies, both the human 7sk and U6 promoters can be used with A and G at the +1 position, and RNAs with the expected length were obtained from them.

The transcriptional start site of the T7 promoter is known to be affected by the first nucleotide downstream of the promoter and several results suggest that at least one Guanine should be used for proper expression. However, it is not consistent in the literature that the T7 promoter requires one, two or 3 guanines.

The U6T7 hybrid promoter was created by modifying the U6 promoter by replacing the end of the promoter with the T7 promoter.

According to the article introducing the hybrid promoter, it is recommended to use at least one G nucleotide. To the best of our knowledge, no result for the +1 position has been published for bacterial J23119 promoters.

Although the U6 and 7sk promoters have been shown to be efficient in expressing RNAs with an A at position +1, in practice targets for SpCas9 gRNAs are generally selected in which the base corresponding to the 5' first base of the spacer is G, which narrows the number of the potential targets. As a solution, various gRNA modifications are used, which in turn can affect the activity of nucleases.

These +1 position studies were generally not performed with next-generation sequencing (NGS), which currently provides the most accurate picture of the length and amount of small RNAs, but with other less accurate methods and the results obtained with NGS so far used a very small sample size.

With regard to Cas12a nucleases we were looking for answers to their ability to effectively induce HR repair in N2A cells, and to their efficiency compared to the most commonly used Cas9 nucleases. We compared the efficacy of As and LbCas12a nucleases with the activity of Sp- and 3 other commonly used Cas9 nucleases. To examine the cleavage activity of nucleases, an assay called GFxFP was optimized from a previously published assay in which a background fluorescent GFP signal was obtained without cleavage of the nuclease. We successfully modified the reporter assay to reduce this non-cleaved GFP background. Using the assay, we wanted to exclude differences in activity caused by different genomic positions due to different PAM sequences. Based on our results, although SpCas9 proved to be more effective: it elicited a higher rate of HR improvement in cells than Cas12a, but compared to other Cas9s, Cas12a nucleases performed to a similar extent at some targets. We tested the suitability of Cas12a nucleases for genome editing not only with a reporter assay but also with genomic targets.



Our results suggest that they are also suitable for inducing HR repair by targeting genomic targets. In our studies of Cas12a nucleases, we used another mammalian cell line (N2a) than Zetsche et al., and demonstrated that Cas12a nucleases are suitable for genome engineering, and while they only tested their activity using the NHEJ repair pathway, we induced HR repair in the genome and we targeted the GFxFP plasmid as well. Based on our results, Cas12a nucleases may provide a useful alternative in the range of genome editing tools, which is consistent with the fact that in recent years, several groups have also shown their suitability for gene editing in a variety of organisms.

In connection with the eSpCas9 and SpCas9HF1 variants with increased specificity, we wanted to know how their activity is affected by the different gRNA modifications used when the base corresponding to the first transcribed base of gRNA in the target is not guanine. Of these modifications, the replacement of the last nucleotide at the 5' end of the gRNA with G, or leaving the spacer unchanged, and the shortening of the 5' end of the gRNAs were tested. Their effects were tested in an EGFP-containing mouse neuroblastoma cell line with the efficacy of EGFP cleavage. Based on our results, while the activity of wild-type SpCas9 is only slightly affected by these modifications, the activity of the high-specificity Cas9 variants is greatly reduced to varying degrees. Thus, based on our results, these nuclease variants cannot be routinely used with such modifications. Several solutions are possible to resolve this, one of which is to modify the nuclease variants with additional mutations that better tolerate the modifications of the gRNA. In order to expand the range of targets for nucleases with increased specificity, there is a need for a more detailed study of promoters suitable for small RNA expression. This could be used to find other sequences with which these nucleases could be used without gRNA modifications.

In most of the work presented in my dissertation, we sought to answer the question of how bases located downstream of the promoters commonly used to express gRNAs are to affect RNA expression and the transcription start positions. In our experiments we performed a much more comprehensive examination of the selected promoters than others before. The role of not only the +1 position but also the 3 additional positions downstream of it was examined for RNA amounts and transcription initiation positions. In our study, plasmid libraries were generated for each promoter and a unique method for sequencing gRNAs was used to sequence RNAs expressed from plasmid libraries.

Our experiments were performed with a large number of different RNA sequences, thanks to randomized barcodes built into the RNA coding region, in order to avoid external effects such as sequence-specific RNA degradation in addition to the 4 base sequence differences studied. Our studies of promoters were performed with high coverage per 4 basepair long sequences using next generation sequencing which is a state-of-the-art technique.

Based on our NGS results, it is true for all investigated promoters that the nucleotides of positions downstream of the promoter can affect the amount of RNA generated and the exact location of transcription initiation, and thus the length of the RNA. So far, only the +1 position has been systematically examined, also only for the T7, U6, and 7sk promoters. Based on our results, not only the +1 position but also the positions downstream of it are important for RNA expression. Thus, there may be orders of magnitude differences in RNA amounts depending on the additional positions, even though the nucleotide at position +1 is the same. It has been found for all promoters tested, that the starting position of transcription may also be affected by a sequence downstream of the promoter. For each promoter, we found that if pyrimidine is in the first position downstream of the promoter, the +1 position will be the starting position in a much smaller proportion, and in this case the transcription will not start from a fixed position. In the case of pyrimidine in +1, the starting position rate of each promoter was greatly influenced not only by the first nucleotide following the promoter but also by the other subsequent nucleotides.

We found that there was no significant difference in the amount of RNA expressed from the human U6 promoter between +1A and +1G, which is consistent with publications that the U6 promoter can also be used with an A. Examination of transcripts transcribed from the U6 promoter showed that if there is an A or G nucleotide at position +1, the starting position of transcription is in almost all cases the +1, which is not much affected by + 2-4 positions.

We found that the human 7sk promoter can also be used not only with +1 guanine but also with adenine for small RNA expression. We have shown that the presence of purine at the +1 position causes accurate transcription initiation at the +1 position, and the additional +2- +4 positions does not change this significantly. +1G produces significantly more RNA than + 1A. We have found that there is no significant difference in the amount of RNA produced from the U6 and 7sk promoters in HEK-293 cells, so the 7sk may be suitable for gRNA expression in parallel or even in place of the U6 promoter.

According to our measurements, the T7 promoter prefers guanine at the +1 position, similar to the previous data. However, there are also sequences starting with adenine from which the same amount of RNAs can be produced like with +1G, and transcription also largely starts from +1. According to our results, GG and GGG sequences generate half and a quarter amount of transcripts, respectively, than those initiating with GA.

Based on our results, the modified U6T7 hybrid promoter, although useful with T7 polymerase, should be considered for mammalian expression. Although we found that there was no significant difference in the amount of RNA generated from the U6T7, 7sk, and U6 promoters, but RNAs from U6T7 promoter are very heterogeneous in length, even from +1purine.

We examined the effect of the sequence downstream of the J23119 promoter on transcription and found that for +1purine, the site of transcription initiation is almost always the +1 position, which is not significantly affected by +2-4. positions. Furthermore, we have shown that most RNA is generated when purine is present in +1.

Our results related to promoters can be significant for several reasons.

For each of the promoters from 3 different organisms we found that not only the first but also the 3 additional positions downstream of the promoter can affect the transcription start site, and thus the length of the RNAs as well as the expression. This may also be interesting in basic research and raises the need for other promoters and additional positions.

Based on our results, higher RNA levels and more accurate lengths of RNAs can be obtained using the sequences provided with the promoters studied. This can be extremely useful for various applications such as RNA interference or CRISPR. Based on our results, the range of targets that can be selected for CRISPR may be expanded due to the finding of downstream-located sequences from promoters that were not previously known to be as suitable for gRNA expression as previously known variations. In addition, our results- together with the results from the examination of additional positions- could be suitable for the further development of the current CRISPR target prediction programs.

The results presented in my dissertation can contribute to a more efficient application of the CRISPR/Cas system.

## 12. Publikációs lista

A disszertáció témájához közvetlenül kapcsolódó közlemények:

1. Cpf1 nucleases demonstrate robust activity to induce DNA modification by exploiting homology directed repair pathways in mammalian cells

Toth, E; **Weinhardt, N\***; Bencsura, P; Huszar, K; Kulcsar, P I; Talas, A; Fodor, E; Welker, E  
*BIOLOGY DIRECT* 11 Paper: 46, 14 p. (2016) **IF:3,472** \*megosztott első szerző

2. Crossing enhanced and high fidelity SpCas9 nucleases to optimize specificity and cleavage

Kulcsar, PI; Talas, A; Huszar, K; Ligeti, Z; Toth, E; **Weinhardt, N**; Fodor, E; Welker, E  
*GENOME BIOLOGY* 18 Paper: 190, 17 p. (2017) **IF:13,214**

3. Nucleotides downstream of the promoter affect start site and RNA level in promoters used for gRNA expression

(Az eredmények jelenleg kézirat formájában vannak, melynek a jelölt az első szerzője)

**A disszertáció témájához közvetlenül kapcsolódó közlemények kumulatív hatástényezője: 16,686**

A disszertáció témájához közvetve kapcsolódó közlemények:

4. Mb- and FnCpf1 nucleases are active in mammalian cells: activities and PAM preferences of four wild-type Cpf1 nucleases and of their altered PAM specificity variants.

Toth, E; Czene, B C; Kulcsar, P I; Krausz, S L; Talas, A; Nyeste, A; Varga, E; Huszar, K; **Weinhardt, N**; Ligeti, Z; Borsy A; Fodor, E; Welker, E  
*NUCLEIC ACIDS RESEARCH* 46: 19 pp. 10272-10285., 14 p. (2018) **IF:11,147**

5. A convenient method to pre-screen candidate guide RNAs for CRISPR/Cas9 gene editing by NHEJ-mediated integration of a 'self-cleaving' GFP-expression plasmid

Talas, A; Kulcsar, P I; **Weinhardt, N**; Borsy, A; Toth, E; Szebenyi, K; Krausz, S L; Huszar, K; Vida, I; Sturm, A; Gordos, B; Hoffmann, O I; Bencsura, P; Nyeste, A; Ligeti, Z; Fodor, E ; Welker, E  
*DNA RESEARCH* 24: 6 pp. 609-621., 13 p. (2017) **IF:5,415**

A disszertáció témájához nem kapcsolódó közlemények:

6. Highly efficient RNAi and Cas9-based auto-cloning systems for C. elegans research.

Sturm, A; Saskoi, E; Kovacs, T; **Weinhardt, N**; Vellai, T  
*NUCLEIC ACIDS RESEARCH* 46: 17 Paper: e105, 13 p. (2018) **IF:11,147**

**MTMT azonosító: 10054747**

**Összes publikáció kumulatív hatástényezője (IF): 44,395**

**Összes független hivatkozások száma: 97**

**Hirsch index: 5**

## 13. Függelék

U6 promóter eredményeihez kapcsolódó p értékek

(0,000 jelentése <0,001)

24. ábra

A	C	0,000
G	1,000	
T	0,000	
C	A	0,000
G	0,000	
T	0,990	
G	A	1,000
C	0,000	
T	0,000	
T	A	0,000
C	0,990	
G	0,000	

25. ábra

AA	AC	1,000	CA	AA	0,510	GA	AA	1,000	TA	AA	1,000
AG	0,000	AC	0,542	AC	1,000	AC	1,000	AC	1,000		
AT	0,511	AG	0,000	AG	0,000	AG	0,000	AG	0,000		
CA	0,510	AT	0,000	AT	0,732	AT	0,523	AT	0,523		
CC	0,000	CC	0,000	CA	0,303	CA	0,498	CA	0,498		
CG	1,000	CG	0,747	CC	0,000	CC	0,000	CC	0,000		
CT	0,000	CT	0,000	CG	1,000	CG	1,000	CG	1,000		
GA	1,000	GA	0,303	CT	0,000	CT	0,000	CT	0,000		
GC	0,789	GC	1,000	GC	0,577	GC	1,000	GC	1,000		
GG	0,000	GG	0,000	GG	0,000	GG	0,779	GG	0,779		
GT	0,003	GT	0,000	GT	0,008	GT	0,000	GT	0,000		
TA	1,000	TA	0,498	TA	1,000	TA	0,003	TA	0,003		
TC	0,000	TC	0,000	TC	0,000	TC	0,000	TC	0,000		
TG	1,000	TG	0,350	TG	1,000	TG	1,000	TG	1,000		
TT	0,000	TT	0,000	TT	0,000	TT	0,000	TT	0,000		
AC	AA	1,000	CC	AA	0,000	GC	AA	0,789	TC	AA	0,000
AG	0,000	AC	0,000	AC	0,815	AC	0,000	AC	0,000		
AT	0,479	AG	0,000	AG	0,000	AG	0,000	AG	0,000		
CA	0,542	AT	0,000	AT	0,002	AT	0,000	AT	0,000		
CC	0,000	CA	0,000	CA	1,000	CA	0,000	CA	0,000		
CG	1,000	CG	0,000	CC	0,000	CC	0,000	CC	0,999		
CT	0,000	CT	0,999	CG	0,938	CG	0,000	CG	0,000		
GA	1,000	GA	0,000	CT	0,000	CT	0,639	CT	0,639		
GC	0,815	GC	0,000	GA	0,577	GA	0,000	GA	0,000		
GG	0,000	GG	0,000	GG	0,000	GG	0,000	GG	0,000		
GT	0,002	GT	0,000	GT	0,000	GT	0,000	GT	0,000		
TA	1,000	TA	0,000	TA	0,779	TA	0,000	TA	0,000		
TC	0,000	TC	0,999	TC	0,000	TC	0,000	TC	0,000		
TG	1,000	TG	0,000	TG	0,634	TG	0,000	TG	0,000		
TT	0,000	TT	1,000	TT	0,000	TT	0,000	TT	1,000		
AG	AA	0,000	CG	AA	1,000	GG	AA	0,000	TG	AA	1,000
AC	0,000	AC	1,000	AC	0,000	AC	0,000	AC	1,000		
AT	0,000	AG	0,000	AG	1,000	AG	0,000	AG	0,000		
CA	0,000	AT	0,289	AT	0,000	AT	0,679	AT	0,679		
CC	0,000	CA	0,747	CA	0,000	CA	0,350	CA	0,350		
CG	0,000	CC	0,000	CC	0,000	CC	0,000	CC	0,000		
CT	0,000	CT	0,000	CG	0,000	CG	1,000	CG	1,000		
GA	0,000	GA	1,000	CT	0,000	CT	0,000	CT	0,000		
GC	0,000	GC	0,938	GA	0,000	GA	1,000	GA	1,000		
GG	1,000	GG	0,000	GC	0,000	GC	0,634	GC	0,634		
GT	0,000	GT	0,001	GT	0,000	GT	0,000	GT	0,000		
TA	0,000	TA	1,000	TA	0,000	TA	0,006	TA	0,006		
TC	0,000	TC	0,000	TC	0,000	TC	1,000	TC	1,000		
TG	0,000	TG	1,000	TG	0,000	TG	0,000	TG	0,000		
TT	0,000	TT	0,000	TT	0,000	TT	0,000	TT	0,000		
AT	AA	0,511	CT	AA	0,000	GT	AA	0,003	TT	AA	0,000
AC	0,479	AC	0,000	AC	0,002	AC	0,000	AC	0,000		
AG	0,000	AG	0,000	AG	0,000	AG	0,000	AG	0,000		
CA	0,000	AT	0,000	AT	0,855	AT	0,000	AT	0,000		
CC	0,000	CA	0,000	CA	0,000	CA	0,000	CA	0,000		
CG	0,289	CC	0,999	CC	0,000	CC	0,000	CC	1,000		
CT	0,000	CG	0,000	CG	0,001	CG	0,000	CG	0,000		
GA	0,732	GA	0,000	CT	0,000	CT	0,983	CT	0,983		
C	0,002	GC	0,000	GA	0,008	GA	0,000	GA	0,000		
GG	0,000	GG	0,000	GC	0,000	GC	0,000	GC	0,000		
GT	0,855	GT	0,000	GG	0,000	GG	0,000	GG	0,000		
TA	0,523	TA	0,000	TA	0,003	TA	0,000	TA	0,000		
TC	0,000	TC	0,639	TC	0,000	TC	0,000	TC	0,000		
TG	0,679	TG	0,000	TG	0,006	TG	1,000	TG	1,000		
TT	0,000	TT	0,983	TT	0,000	TT	0,000	TT	0,000		

## 26. ábra

AAA	AAC	0,619	AAG	AAA	0,431	ACA	AAA	0,604	ACG	AAA	0,657	AGA	AAA	0,000	AGG	AAA	0,002	ATA	AAA	0,262	ATG	AAA	0,132
	AAG	0,431		AAC	0,973		AAC	1,000		AAC	1,000		AAC	0,000		AAC	0,013		AAC	0,827		AAC	0,756
	AAT	0,122		AAT	1,000		AAG	1,000		AAG	1,000		AAG	0,823		AAG	0,159		AAG	1,000		AAG	1,000
	ACA	0,604		ACA	1,000		AAT	1,000		AAT	1,000		AAT	0,049		AAT	0,040		AAT	1,000		AAT	1,000
	ACC	0,887		ACC	0,980		ACC	1,000		ACA	1,000		ACA	0,019		ACA	0,018		ACA	0,987		ACA	0,996
	ACG	0,657		ACG	1,000		ACG	1,000		ACC	1,000		ACC	0,002		ACC	0,010		ACC	0,851		ACC	0,829
	ACT	0,937		ACT	0,991		ACT	1,000		ACT	1,000		ACG	0,242		ACG	0,041		ACG	1,000		ACG	1,000
	AGA	0,000		AGA	0,823		AGA	0,019		AGA	0,242		ACT	0,008		ACT	0,010		ACT	0,903		ACT	0,912
	AGC	0,394		AGC	0,994		AGC	0,845		AGC	0,935		AGC	1,000		AGA	0,654		AGA	0,988		AGA	0,550
	AGG	0,002		AGG	0,159		AGG	0,018		AGG	0,041		AGG	0,654		AGC	1,000		AGC	1,000		AGC	0,992
	AGT	0,000		AGT	0,095		AGT	0,001		AGT	0,010		AGT	0,367		AGT	1,000		AGG	0,330		AGG	0,097
	ATA	0,262		ATA	1,000		ATA	0,987		ATA	1,000		ATA	0,988		ATA	0,330		AGT	0,249		AGT	0,020
	ATC	0,067		ATC	1,000		ATC	0,999		ATC	1,000		ATC	0,049		ATC	0,048		ATC	1,000		ATA	1,000
	ATG	0,132		ATG	1,000		ATG	0,996		ATG	1,000		ATG	0,550		ATG	0,097		ATG	1,000		ATC	1,000
	ATT	0,323		ATT	1,000		ATT	1,000		ATT	1,000		ATT	0,168		ATT	0,042		ATT	1,000		ATT	1,000
	GAA	0,513		GAA	0,995		GAA	1,000		GAA	1,000		GAA	0,001		GAA	0,015		GAA	0,914		GAA	0,905
	GAC	0,898		GAC	0,898		GAC	0,999		GAC	0,996		GAC	0,000		GAC	0,009		GAC	0,686		GAC	0,545
	GAG	0,019		GAG	1,000		GAG	0,865		GAG	0,999		GAG	0,303		GAG	0,095		GAG	1,000		GAG	1,000
	GAT	0,929		GAT	0,997		GAT	1,000		GAT	1,000		GAT	0,019		GAT	0,012		GAT	0,943		GAT	0,961
	GCA	0,568		GCA	0,986		GCA	1,000		GCA	1,000		GCA	0,001		GCA	0,014		GCA	0,872		GCA	0,834
	GCC	1,000		GCC	0,387		GCC	0,574		GCC	0,604		GCC	0,001		GCC	0,001		GCC	0,231		GCC	0,133
	GCG	0,309		GCG	1,000		GCG	1,000		GCG	1,000		GCG	0,002		GCG	0,020		GCG	0,971		GCG	0,981
	GCT	1,000		GCT	0,359		GCT	0,490		GCT	0,557		GCT	0,000		GCT	0,002		GCT	0,214		GCT	0,101
	GGA	0,001		GGA	0,851		GGA	0,020		GGA	0,264		GGA	1,000		GGA	0,497		GGA	0,994		GGA	0,562
	GGC	0,052		GGC	0,989		GGC	0,466		GGC	0,768		GGC	1,000		GGC	0,941		GGC	1,000		GGC	0,971
	GGG	0,005		GGG	0,085		GGG	0,025		GGG	0,036		GGG	0,347		GGG	1,000		GGG	0,152		GGG	0,071
	GGT	0,003		GGT	0,531		GGT	0,051		GGT	0,154		GGT	0,998		GGT	1,000		GGT	0,821		GGT	0,356
	GTA	0,000		GTA	0,675		GTA	0,010		GTA	0,157		GTA	1,000		GTA	0,659		GTA	0,948		GTA	0,333
	GTC	0,142		GTC	1,000		GTC	1,000		GTC	1,000		GTC	0,005		GTC	0,029		GTC	0,997		GTC	1,000
	GTG	0,000		GTG	0,916		GTG	0,033		GTG	0,344		GTG	1,000		GTG	0,524		GTG	0,998		GTG	0,722
	GTT	0,198		GTT	1,000		GTT	0,994		GTT	1,000		GTT	0,842		GTT	0,162		GTT	1,000		GTT	1,000
AAC	AAA	0,619	AAT	AAA	0,122	ACC	AAA	0,887	ACT	AAA	0,937	AGC	AAA	0,394	AGT	AAA	0,000	ATC	AAA	0,067	ATT	AAA	0,323
	AAG	0,973		AAC	0,845		AAC	1,000		AAC	1,000		AAC	0,721		AAC	0,000		AAC	0,567		AAC	0,989
	AAT	0,845		AAG	1,000		AAG	0,980		AAG	0,991		AAG	0,994		AAG	0,095		AAG	1,000		AAG	1,000
	ACA	1,000		ACA	1,000		AAT	0,936		AAT	0,982		AAT	0,947		AAT	0,002		AAT	1,000		AAT	1,000
	ACC	1,000		ACC	0,936		ACA	1,000		ACA	1,000		ACA	0,845		ACA	0,001		ACA	0,999		ACA	1,000
	ACG	1,000		ACG	1,000		ACG	1,000		ACC	1,000		ACC	0,717		ACC	0,000		ACC	0,783		ACC	0,995
	ACT	1,000		ACT	0,982		ACT	1,000		ACG	1,000		ACG	0,935		ACG	0,010		ACG	1,000		ACG	1,000
	AGA	0,000		AGA	0,049		AGA	0,002		AGA	0,008		ACT	0,741		ACT	0,000		ACT	0,917		ACT	0,999
	AGC	0,721		AGC	0,947		AGC	0,717		AGC	0,741		AGA	1,000		AGA	0,367		AGA	0,049		AGA	0,168
	AGG	0,013		AGG	0,040		AGG	0,010		AGG	0,010		AGG	1,000		AGC	1,000		AGC	0,961		AGC	0,950
	AGT	0,000		AGT	0,002		AGT	0,000		AGT	0,000		AGT	1,000		AGG	1,000		AGG	0,048		AGG	0,042
	ATA	0,827		ATA	1,000		ATA	0,851		ATA	0,903		ATA	1,000		ATA	0,249		AGT	0,002		AGT	0,005
	ATC	0,567		ATC	1,000		ATC	0,783		ATC	0,917		ATC	0,961		ATC	0,002		ATA	1,000		ATA	1,000
	ATG	0,756		ATG	1,000		ATG	0,829		ATG	0,912		ATG	0,992		ATG	0,020		ATG	1,000		ATC	1,000
	ATT	0,989		ATT	1,000		ATT	0,995		ATT	0,999		ATT	0,950		ATT	0,005		ATT	1,000		ATG	1,000
	GAA	1,000		GAA	0,981		GAA	1,000		GAA	1,000		GAA	0,773		GAA	0,000		GAA	0,869		GAA	1,000
	GAC	1,000		GAC	0,560		GAC	1,000		GAC	1,000		GAC	0,645		GAC	0,000		GAC	0,288		GAC	0,904
	GAG	0,161		GAG	1,000		GAG	0,317		GAG	0,525		GAG	0,994		GAG	0,008		GAG	1,000		GAG	1,000
	GAT	1,000		GAT	0,996		GAT	1,000		GAT	1,000		GAT	0,772		GAT	0,001		GAT	0,972		GAT	1,000
	GCA	1,000		GCA	0,932		GCA	1,000		GCA	1,000		GCA	0,746		GCA	0,000		GCA	0,725		GCA	0,997
	GCC	0,623		GCC	0,154		GCC	0,842		GCC	0,896		GCC	0,356		GCC	0,000		GCC	0,099		GCC	0,313
	GCG	1,000		GCG	1,000		GCG	1,000		GCG	1,000		GCG	0,830		GCG	0,001		GCG	0,987		GCG	1,000
	GCT	0,502		GCT	0,094		GCT	0,786		GCT	0,864		GCT	0,358		GCT	0,000		GCT	0,053		GCT	0,252
	GGA	0,000		GGA	0,038		GGA	0,002		GGA	0,011		GGA	1,000		GGA	0,170		GGA	0,028		GGA	0,176
	GGC	0,250		GGC	0,742		GGC	0,255		GGC	0,296		GGC	1,000		GGC	0,946		GGC	0,790		GGC	0,791
	GGG	0,020		GGG	0,044		GGG	0,017		GGG	0,017		GGG	0,989		GGG	0,981		GGG	0,050		GGG	0,041
	GGT	0,026		GGT	0,121		GGT	0,022		GGT	0,026		GGT	1,000		GGT	1,000		GGT	0,141		GGT	0,149
	GTA	0,000		GTA	0,014		GTA	0,001		GTA	0,005		GTA	1,000		GTA	0,325		GTA	0,009		GTA	0,090
	GTC	0,902		GTC	1,000		GTC	0,979		GTC	0,997		GTC	0,893		GTC	0,001		GTC	1,000		GTC	1,000
	GTG	0,001		GTG	0,090		GTG	0,003		GTG	0,014		GTG	1,000		GTG	0,215		GTG	0,092		GTG	0,262
	GTT	0,815		GTT	1,000		GTT	0,856		GTT	0,918		GTT	0,997		GTT	0,072		GTT	1,000		GTT	1,000

## 26. ábra

GAA	AAA	0,513	GAG	AAA	0,019	GCA	AAA	0,568	GCG	AAA	0,309	GGA	AAA	0,001	GGG	AAA	0,005	GTA	AAA	0,000	GTG	AAA	0,000
AAC	1,000		AAC	0,161		AAC	1,000		AAC	1,000		AAC	0,000		AAC	0,020		AAC	0,000		AAC	0,001	
AAG	0,995		AAG	1,000		AAG	0,986		AAG	1,000		AAG	0,851		AAG	0,085		AAG	0,675		AAG	0,916	
AAT	0,981		AAT	1,000		AAT	0,932		AAT	1,000		AAT	0,038		AAT	0,044		AAT	0,014		AAT	0,090	
ACA	1,000		ACA	0,865		ACA	1,000		ACA	1,000		ACA	0,020		ACA	0,025		ACA	0,010		ACA	0,033	
ACC	1,000		ACC	0,317		ACC	1,000		ACC	1,000		ACC	0,002		ACC	0,017		ACC	0,001		ACC	0,003	
ACG	1,000		ACG	0,999		ACG	1,000		ACG	1,000		ACG	0,264		ACG	0,036		ACG	0,157		ACG	0,344	
ACT	1,000		ACT	0,525		ACT	1,000		ACT	1,000		ACT	0,011		ACT	0,017		ACT	0,005		ACT	0,014	
AGA	0,001		AGA	0,303		AGA	0,001		AGA	0,002		AGA	1,000		AGA	0,347		AGA	1,000		AGA	1,000	
AGC	0,773		AGC	0,994		AGC	0,746		AGC	0,830		AGC	1,000		AGC	0,989		AGC	1,000		AGC	1,000	
AGG	0,015		AGG	0,095		AGG	0,014		AGG	0,020		AGG	0,497		AGG	1,000		AGG	0,659		AGG	0,524	
AGT	0,000		AGT	0,008		AGT	0,000		AGT	0,001		AGT	0,170		AGT	0,981		AGT	0,325		AGT	0,215	
ATA	0,914		ATA	1,000		ATA	0,872		ATA	0,971		ATA	0,994		ATA	0,152		ATA	0,948		ATA	0,998	
ATC	0,869		ATC	1,000		ATC	0,725		ATC	0,987		ATC	0,028		ATC	0,050		ATC	0,009		ATC	0,092	
ATG	0,905		ATG	1,000		ATG	0,834		ATG	0,981		ATG	0,562		ATG	0,071		ATG	0,333		ATG	0,722	
ATT	1,000		ATT	1,000		ATT	0,997		ATT	1,000		ATT	0,176		ATT	0,041		ATT	0,090		ATT	0,262	
GAC	1,000		GAA	0,348		GAA	1,000		GAA	1,000		GAA	0,000		GAA	0,023		GAA	0,000		GAA	0,002	
GAG	0,348		GAC	0,071		GAC	1,000		GAC	0,957		GAC	0,000		GAC	0,016		GAC	0,000		GAC	0,000	
GAT	1,000		GAT	0,687		GAG	0,235		GAG	0,592		GAG	0,240		GAG	0,080		GAG	0,086		GAG	0,497	
GCA	1,000		GCA	0,235		GAT	1,000		GAT	1,000		GAT	0,024		GAT	0,018		GAT	0,013		GAT	0,031	
GCC	0,528		GCC	0,035		GCC	0,577		GCA	1,000		GCA	0,000		GCA	0,022		GCA	0,000		GCA	0,001	
GCG	1,000		GCG	0,592		GCG	1,000		GCC	0,353		GCC	0,002		GCC	0,004		GCC	0,001		GCC	0,002	
GCT	0,408		GCT	0,016		GCT	0,456		GCT	0,242		GCG	0,000		GCG	0,028		GCG	0,000		GCG	0,003	
GGA	0,000		GGA	0,240		GGA	0,000		GGA	0,000		GCT	0,001		GCT	0,005		GCT	0,000		GCT	0,000	
GGC	0,314		GGC	0,969		GGC	0,278		GGC	0,404		GGC	1,000		GGA	0,282		GGA	1,000		GGA	1,000	
GGG	0,023		GGG	0,080		GGG	0,022		GGG	0,028		GGG	0,282		GGC	0,565		GGC	1,000		GGC	1,000	
GGT	0,033		GGT	0,313		GGT	0,029		GGT	0,046		GGT	0,973		GGT	0,887		GGG	0,363		GGG	0,284	
GTA	0,000		GTA	0,086		GTA	0,000		GTA	0,000		GTA	1,000		GTA	0,363		GGT	0,998		GGT	0,982	
GTC	0,998		GTC	0,895		GTC	0,978		GTC	1,000		GTC	0,001		GTC	0,036		GTC	0,000		GTA	1,000	
GTG	0,002		GTG	0,497		GTG	0,001		GTG	0,003		GTG	1,000		GTG	0,284		GTG	1,000		GTC	0,009	
GTT	0,922		GTT	1,000		GTT	0,871		GTT	0,981		GTT	0,865		GTT	0,094		GTT	0,665		GTT	0,937	
GAC	AAA	0,898	GAT	AAA	0,929	GCC	AAA	1,000	GCT	AAA	1,000	GCG	AAA	0,052	GGT	AAA	0,003	GTC	AAA	0,142	GTT	AAA	0,198
AAC	1,000		AAC	1,000		AAC	0,623		AAC	0,502		AAC	0,250		AAC	0,026		AAC	0,902		AAC	0,815	
AAG	0,898		AAG	0,997		AAG	0,387		AAG	0,359		AAG	0,989		AAG	0,531		AAG	1,000		AAG	1,000	
AAT	0,560		AAT	0,996		AAT	0,154		AAT	0,094		AAT	0,742		AAT	0,121		AAT	1,000		AAT	1,000	
ACA	0,999		ACA	1,000		ACA	0,574		ACA	0,490		ACA	0,466		ACA	0,051		ACA	1,000		ACA	0,994	
ACC	1,000		ACC	1,000		ACC	0,842		ACC	0,786		ACC	0,255		ACC	0,022		ACC	0,979		ACC	0,856	
ACG	0,996		ACG	1,000		ACG	0,604		ACG	0,557		ACG	0,768		ACG	0,154		ACG	1,000		ACG	1,000	
ACT	1,000		ACT	1,000		ACT	0,896		ACT	0,864		ACT	0,296		ACT	0,026		ACT	0,997		ACT	0,918	
AGA	0,000		AGA	0,019		AGA	0,001		AGA	0,000		AGA	1,000		AGA	0,998		AGA	0,005		AGA	0,842	
AGC	0,645		AGC	0,772		AGC	0,356		AGC	0,358		AGC	1,000		AGC	1,000		AGC	0,893		AGC	0,997	
AGG	0,009		AGG	0,012		AGG	0,001		AGG	0,002		AGG	0,941		AGG	1,000		AGG	0,029		AGG	0,162	
AGT	0,000		AGT	0,001		AGT	0,000		AGT	0,000		AGT	0,946		AGT	1,000		AGT	0,001		AGT	0,072	
ATA	0,686		ATA	0,943		ATA	0,231		ATA	0,214		ATA	1,000		ATA	0,821		ATA	0,997		ATA	1,000	
ATC	0,288		ATC	0,972		ATC	0,099		ATC	0,053		ATC	0,790		ATC	0,141		ATC	1,000		ATC	1,000	
ATG	0,545		ATG	0,961		ATG	0,133		ATG	0,101		ATG	0,971		ATG	0,356		ATG	1,000		ATG	1,000	
ATT	0,904		ATT	1,000		ATT	0,313		ATT	0,252		ATT	0,791		ATT	0,149		ATT	1,000		ATT	1,000	
GAA	1,000		GAA	1,000		GAA	0,528		GAA	0,408		GAA	0,314		GAA	0,033		GAA	0,998		GAA	0,922	
GAG	0,071		GAC	1,000		GAC	0,861		GAC	0,794		GAC	0,181		GAC	0,018		GAC	0,548		GAC	0,642	
GAT	1,000		GAG	0,687		GAG	0,035		GAG	0,016		GAG	0,969		GAG	0,313		GAG	0,895		GAG	1,000	
GCA	1,000		GCA	1,000		GAT	0,886		GAT	0,855		GAT	0,350		GAT	0,033		GAT	1,000		GAT	0,960	
GCC	0,861		GCC	0,886		GCA	0,577		GCA	0,456		GCA	0,278		GCA	0,029		GCA	0,978		GCA	0,871	
GCG	0,957		GCG	1,000		GCG	0,353		GCC	1,000		GCC	0,043		GCC	0,003		GCC	0,194		GCC	0,181	
GCT	0,794		GCT	0,855		GCT	1,000		GCG	0,242		GCG	0,404		GCG	0,046		GCG	1,000		GCG	0,981	
GGA	0,000		GGA	0,024		GGA	0,002		GGA	0,001		GCT	0,042		GCT	0,002		GCT	0,114		GCT	0,156	
GGC	0,181		GGC	0,350		GGC	0,043		GGC	0,042		GGA	1,000		GGA	0,973		GGA	0,001		GGA	0,865	
GGG	0,016		GGG	0,018		GGG	0,004		GGG	0,005		GGG	0,565		GGG	1,000		GGC	0,539		GGC	0,995	
GGT	0,018		GGT	0,033		GGT	0,003		GGT	0,002		GGT	1,000		GGG	0,887		GGG	0,036		GGG	0,094	
GTA	0,000		GTA	0,013		GTA	0,001		GTA	0,000		GTA	1,000		GTA	0,998		GGT	0,071		GGT	0,551	
GTC	0,548		GTC	1,000		GTC	0,194		GTC	0,114		GTC	0,539		GTC	0,071		GTA	0,000		GTA	0,665	
GTG	0,000		GTG	0,031		GTG	0,002		GTG	0,000		GTG	1,000		GTG	0,982		GTG	0,009		GTC	0,999	
GTT	0,642		GTT	0,960		GTT	0,181		GTT	0,156		GTT	0,995		GTT	0,551		GTT	0,999		GTG	0,937	

7sk promóter eredményeihez kapcsolódó p értékek

(0,000 jelentése <0,001)

32. ábra

A	C	0,000
G	0,000	
T	0,000	
C	A	0,000
G	0,000	
T	1,000	
G	A	0,000
C	0,000	
T	0,000	
T	A	0,000
C	1,000	
G	0,000	

33. ábra

AA	AC	1,000	CA	AA	0,000	GA	AA	0,001	TA	AA	0,031
AG	0,000		AC	0,000		AC	0,007		AC	0,007	
AT	0,984		AG	0,000		AG	0,000		AG	0,000	
CA	0,000		AT	0,000		AT	0,150		AT	0,000	
CC	0,000		CC	0,006		CA	0,000		CA	0,998	
CG	1,000		CG	0,004		CC	0,000		CC	0,000	
CT	0,000		CT	0,000		CG	0,000		CG	0,164	
GA	0,001		GA	0,000		CT	0,000		CT	0,000	
GC	1,000		GC	0,000		GC	0,004		GA	0,000	
GG	0,000		GG	0,000		GG	0,000		GC	0,012	
GT	0,000		GT	0,000		GT	0,000		GG	0,000	
TA	0,031		TA	0,998		TA	0,000		GT	0,000	
TC	0,000		TC	0,000		TC	0,000		TC	0,000	
TG	1,000		TG	0,000		TG	0,006		TG	0,008	
TT	0,000		TT	0,000		TT	0,000		TT	0,000	
AC	AA	1,000	CC	AA	0,000	GC	AA	1,000	TC	AA	0,000
AG	0,000		AC	0,000		AC	1,000		AC	0,000	
AT	1,000		AG	0,000		AG	0,000		AG	0,000	
CA	0,000		AT	0,000		AT	0,998		AT	0,000	
CC	0,000		CA	0,006		CA	0,000		CA	0,000	
CG	1,000		CG	0,000		CC	0,000		CC	0,969	
CT	0,000		CT	0,994		CG	1,000		CG	0,000	
GA	0,007		GA	0,000		CT	0,000		CT	1,000	
GC	1,000		GC	0,000		GA	0,004		GA	0,000	
GG	0,000		GG	0,000		GG	0,000		GC	0,000	
GT	0,000		GT	0,000		GT	0,000		GG	0,000	
TA	0,007		TA	0,000		TA	0,012		GT	0,000	
TC	0,000		TC	0,969		TC	0,000		TA	0,000	
TG	1,000		TG	0,000		TG	1,000		TG	0,000	
TT	0,000		TT	0,525		TT	0,000		TT	1,000	
AG	AA	0,000	CG	AA	1,000	GG	AA	0,000	TG	AA	1,000
AC	0,000		AC	1,000		AC	0,000		AC	1,000	
AT	0,000		AG	0,000		AG	0,000		AG	0,000	
CA	0,000		AT	0,781		AT	0,000		AT	1,000	
CC	0,000		CA	0,004		CA	0,000		CA	0,000	
CG	0,000		CC	0,000		CC	0,000		CC	0,000	
CT	0,000		CT	0,000		CG	0,000		CG	1,000	
GA	0,000		GA	0,000		CT	0,000		CT	0,000	
GC	0,000		GC	1,000		GA	0,000		GA	0,006	
GG	0,000		GG	0,000		GC	0,000		GC	1,000	
GT	1,000		GT	0,000		GT	0,000		GG	0,000	
TA	0,000		TA	0,164		TA	0,000		GT	0,000	
TC	0,000		TC	0,000		TC	0,000		TA	0,008	
TG	0,000		TG	1,000		TG	0,000		TC	0,000	
TT	0,000		TT	0,000		TT	0,000		TT	0,000	
AT	AA	0,984	CT	AA	0,000	GT	AA	0,000	TT	AA	0,000
AC	1,000		AC	0,000		AC	0,000		AC	0,000	
AG	0,000		AG	0,000		AG	1,000		AG	0,000	
CA	0,000		AT	0,000		AT	0,000		AT	0,000	
CC	0,000		CA	0,000		CA	0,000		CA	0,000	
CG	0,781		CC	0,994		CC	0,000		CC	0,525	
CT	0,000		CG	0,000		CG	0,000		CG	0,000	
GA	0,150		GA	0,000		CT	0,000		CT	0,999	
C	0,998		GC	0,000		GA	0,000		GA	0,000	
GG	0,000		GG	0,000		GC	0,000		GC	0,000	
GT	0,000		GT	0,000		GG	0,000		GG	0,000	
TA	0,000		TA	0,000		TA	0,000		GT	0,000	
TC	0,000		TC	1,000		TC	0,000		TA	0,000	
TG	1,000		TG	0,000		TG	0,000		TC	1,000	
TT	0,000		TT	0,999		TT	0,000		TG	0,000	



34. ábra

AAA	AAC	0,087	AAG	AAA	0,008	ACA	AAA	0,005	ACG	AAA	0,399	AGA	AAA	0,001	AGG	AAA	0,000	ATA	AAA	0,106	ATG	AAA	0,013
AAG	0,008		AAC	0,953		AAC	0,456		AAC	1,000		AAC	0,066		AAC	0,000		AAC	0,984		AAC	0,991	
AAT	0,054		AAT	0,839		AAG	1,000		AAG	1,000		AAG	0,344		AAG	0,003		AAG	1,000		AAG	1,000	
ACA	0,005		ACA	1,000		AAT	0,041		AAT	1,000		AAT	0,065		AAT	0,000		AAT	0,974		AAT	0,958	
ACC	0,866		ACC	0,708		ACC	0,421		ACA	0,992		ACA	0,314		ACA	0,000		ACA	1,000		ACA	1,000	
ACG	0,399		ACG	1,000		ACG	0,992		ACC	1,000		ACC	0,030		ACC	0,002		ACC	0,853		ACC	0,820	
ACT	0,889		ACT	0,255		ACT	0,080		ACT	0,999		ACG	0,174		ACG	0,023		ACG	0,999		ACG	1,000	
AGA	0,001		AGA	0,344		AGA	0,314		AGA	0,174		ACT	0,013		ACT	0,000		ACT	0,633		ACT	0,386	
AGC	0,000		AGC	0,106		AGC	0,097		AGC	0,059		AGC	1,000		AGA	1,000		AGA	0,947		AGA	0,323	
AGG	0,000		AGG	0,003		AGG	0,000		AGG	0,023		AGG	1,000		AGC	1,000		AGC	0,796		AGC	0,100	
AGT	0,000		AGT	0,001		AGT	0,000		AGT	0,008		AGT	1,000		AGT	1,000		AGG	0,637		AGG	0,006	
ATA	0,106		ATA	1,000		ATA	1,000		ATA	0,999		ATA	0,947		ATA	0,637		AGT	0,354		AGT	0,001	
ATC	0,011		ATC	1,000		ATC	1,000		ATC	1,000		ATC	0,352		ATC	0,007		ATC	1,000		ATA	1,000	
ATG	0,013		ATG	1,000		ATG	1,000		ATG	1,000		ATG	0,323		ATG	0,006		ATG	1,000		ATC	1,000	
ATT	1,000		ATT	0,003		ATT	0,001		ATT	0,457		ATT	0,002		ATT	0,000		ATT	0,134		ATT	0,007	
GAA	0,002		GAA	1,000		GAA	0,999		GAA	0,918		GAA	0,774		GAA	0,063		GAA	1,000		GAA	1,000	
GAC	0,013		GAC	1,000		GAC	0,809		GAC	1,000		GAC	0,168		GAC	0,000		GAC	1,000		GAC	1,000	
GAG	0,021		GAG	0,876		GAG	0,861		GAG	0,618		GAG	1,000		GAG	1,000		GAG	1,000		GAG	0,856	
GAT	0,042		GAT	1,000		GAT	1,000		GAT	1,000		GAT	0,647		GAT	0,142		GAT	1,000		GAT	1,000	
GCA	0,020		GCA	1,000		GCA	0,999		GCA	1,000		GCA	0,178		GCA	0,000		GCA	1,000		GCA	1,000	
GCC	0,995		GCC	0,952		GCC	0,868		GCC	1,000		GCC	0,122		GCC	0,046		GCC	0,956		GCC	0,975	
GCG	0,017		GCG	1,000		GCG	1,000		GCG	1,000		GCG	0,398		GCG	0,018		GCG	1,000		GCG	1,000	
GCT	1,000		GCT	0,257		GCT	0,143		GCT	0,977		GCT	0,009		GCT	0,001		GCT	0,483		GCT	0,339	
GGA	0,001		GGA	0,056		GGA	0,063		GGA	0,023		GGA	0,985		GGA	0,932		GGA	0,263		GGA	0,051	
GGC	0,001		GGC	0,038		GGC	0,044		GGC	0,015		GGC	0,744		GGC	0,598		GGC	0,116		GGC	0,034	
GGG	0,000		GGG	0,002		GGG	0,004		GGG	0,001		GGG	0,284		GGG	0,140		GGG	0,016		GGG	0,002	
GGT	0,001		GGT	0,034		GGT	0,040		GGT	0,013		GGT	0,882		GGT	0,731		GGT	0,146		GGT	0,030	
GTA	0,000		GTA	0,000		GTA	0,000		GTA	0,000		GTA	0,346		GTA	0,006		GTA	0,021		GTA	0,000	
GTC	0,000		GTC	0,034		GTC	0,000		GTC	0,095		GTC	1,000		GTC	0,794		GTC	0,957		GTC	0,049	
GTG	0,000		GTG	0,000		GTG	0,000		GTG	0,002		GTG	0,950		GTG	0,297		GTG	0,113		GTG	0,000	
GTT	0,030		GTT	0,962		GTT	0,955		GTT	0,768		GTT	1,000		GTT	1,000		GTT	1,000		GTT	0,951	
AAC	AAA	0,087	AAT	AAA	0,054	ACC	AAA	0,866	ACT	AAA	0,889	AGC	AAA	0,000	AGT	AAA	0,000	ATC	AAA	0,011	ATT	AAA	1,000
AAG	0,953		AAC	1,000		AAC	1,000		AAC	0,968		AAC	0,013		AAC	0,000		AAC	0,981		AAC	0,045	
AAT	1,000		AAG	0,839		AAG	0,708		AAG	0,255		AAG	0,106		AAG	0,001		AAG	1,000		AAG	0,003	
ACA	0,456		ACA	0,041		AAT	0,999		AAT	0,856		AAT	0,014		AAT	0,000		AAT	0,927		AAT	0,014	
ACC	1,000		ACC	0,999		ACA	0,421		ACA	0,080		ACA	0,097		ACA	0,000		ACA	1,000		ACA	0,001	
ACG	1,000		ACG	1,000		ACG	1,000		ACC	1,000		ACC	0,006		ACC	0,001		ACC	0,779		ACC	0,922	
ACT	0,968		ACT	0,856		ACT	1,000		ACG	0,999		ACG	0,059		ACG	0,008		ACG	1,000		ACG	0,457	
AGA	0,066		AGA	0,065		AGA	0,030		AGA	0,013		ACT	0,002		ACT	0,000		ACT	0,339		ACT	0,940	
AGC	0,013		AGC	0,014		AGC	0,006		AGC	0,002		AGA	1,000		AGA	1,000		AGA	0,352		AGA	0,002	
AGG	0,000		AGG	0,000		AGG	0,002		AGG	0,000		AGG	1,000		AGC	1,000		AGC	0,112		AGC	0,000	
AGT	0,000		AGT	0,000		AGT	0,001		AGT	0,000		AGT	1,000		AGG	1,000		AGG	0,007		AGG	0,000	
ATA	0,984		ATA	0,974		ATA	0,853		ATA	0,633		ATA	0,796		ATA	0,354		AGT	0,001		AGT	0,000	
ATC	0,981		ATC	0,927		ATC	0,779		ATC	0,339		ATC	0,112		ATC	0,001		ATA	1,000		ATA	0,134	
ATG	0,991		ATG	0,958		ATG	0,820		ATG	0,386		ATG	0,100		ATG	0,001		ATG	1,000		ATC	0,006	
ATT	0,045		ATT	0,014		ATT	0,922		ATT	0,940		ATT	0,000		ATT	0,000		ATT	0,006		ATG	0,007	
GAA	0,379		GAA	0,231		GAA	0,261		GAA	0,050		GAA	0,414		GAA	0,011		GAA	1,000		GAA	0,001	
GAC	0,928		GAC	0,370		GAC	0,728		GAC	0,224		GAC	0,046		GAC	0,000		GAC	1,000		GAC	0,002	
GAG	0,406		GAG	0,382		GAG	0,230		GAG	0,137		GAG	1,000		GAG	1,000		GAG	0,877		GAG	0,030	
GAT	0,990		GAT	0,977		GAT	0,852		GAT	0,543		GAT	0,340		GAT	0,047		GAT	1,000		GAT	0,049	
GCA	1,000		GCA	0,999		GCA	0,936		GCA	0,554		GCA	0,043		GCA	0,000		GCA	1,000		GCA	0,008	
GCC	1,000		GCC	1,000		GCC	1,000		GCC	1,000		GCC	0,059		GCC	0,023		GCC	0,966		GCC	0,999	
GCG	0,990		GCG	0,964		GCG	0,825		GCG	0,425		GCG	0,140		GCG	0,004		GCG	1,000		GCG	0,013	
GCT	0,860		GCT	0,745		GCT	1,000		GCT	1,000		GCT	0,002		GCT	0,000		GCT	0,306		GCT	1,000	
GGA	0,016		GGA	0,019		GGA	0,006		GGA	0,004		GGA	0,993		GGA	0,997		GGA	0,055		GGA	0,001	
GGC	0,014		GGC	0,016		GGC	0,006		GGC	0,005		GGC	0,790		GGC	0,825		GGC	0,036		GGC	0,002	
GGG	0,001		GGG	0,001		GGG	0,000		GGG	0,000		GGG	0,301		GGG	0,297		GGG	0,002		GGG	0,000	
GGT	0,011		GGT	0,013		GGT	0,004		GGT	0,003		GGT	0,917		GGT	0,937		GGT	0,033		GGT	0,001	
GTA	0,000		GTA	0,000		GTA	0,000		GTA	0,000		GTA	0,303		GTA	0,059		GTA	0,000		GTA	0,000	
GTC	0,000		GTC	0,000		GTC	0,009		GTC	0,001		GTC	0,997		GTC	0,143		GTC	0,058		GTC	0,000	
GTG	0,000		GTG	0,000		GTG	0,000		GTG	0,000		GTG	0,972		GTG	0,949		GTG	0,000		GTG	0,000	
GTT	0,542		GTT	0,509		GTT	0,325		GTT	0,195		GTT	1,000		GTT	0,989		GTT	0,962		GTT	0,041	

34. ábra

GAA	AAA	0,002	GAG	AAA	0,021	GCA	AAA	0,020	GCG	AAA	0,017	GGA	AAA	0,001	GGG	AAA	0,000	GTA	AAA	0,000	GTG	AAA	0,000
AAC	0,379	AAC	0,406	AAC	1,000	AAC	0,990	AAC	0,016	AAC	0,001	AAC	0,000	AAC	0,000	AAC	0,000	AAC	0,000	AAC	0,000	AAC	0,000
AAG	1,000	AAG	0,876	AAG	1,000	AAG	1,000	AAG	0,056	AAG	0,002	AAG	0,000	AAG	0,000	AAG	0,000	AAG	0,000	AAG	0,000	AAG	0,000
AAT	0,231	AAT	0,382	AAT	0,999	AAT	0,964	AAT	0,019	AAT	0,001	AAT	0,000	AAT	0,000	AAT	0,000	AAT	0,000	AAT	0,000	AAT	0,000
ACA	0,999	ACA	0,861	ACA	0,999	ACA	1,000	ACA	0,063	ACA	0,004	ACA	0,000	ACA	0,000	ACA	0,000	ACA	0,000	ACA	0,000	ACA	0,000
ACC	0,261	ACC	0,230	ACC	0,936	ACC	0,825	ACC	0,006	ACC	0,000	ACC	0,000	ACC	0,000	ACC	0,000	ACC	0,000	ACC	0,000	ACC	0,000
ACG	0,918	ACG	0,618	ACG	1,000	ACG	1,000	ACG	0,023	ACG	0,001	ACG	0,000	ACG	0,000	ACG	0,000	ACG	0,000	ACG	0,002	ACG	0,002
ACT	0,050	ACT	0,137	ACT	0,554	ACT	0,425	ACT	0,004	ACT	0,000	ACT	0,000	ACT	0,000	ACT	0,000	ACT	0,000	ACT	0,000	ACT	0,000
AGA	0,774	AGA	1,000	AGA	0,178	AGA	0,398	AGA	0,985	AGA	0,284	AGA	0,346	AGA	0,950	AGA	0,303	AGA	0,303	AGA	0,972	AGA	0,972
AGC	0,414	AGC	1,000	AGC	0,043	AGC	0,140	AGC	0,993	AGC	0,301	AGC	0,006	AGC	0,297	AGC	0,059	AGC	0,059	AGC	0,949	AGC	0,949
AGG	0,063	AGG	1,000	AGG	0,000	AGG	0,018	AGG	0,932	AGG	0,140	AGG	0,006	AGG	0,297	AGG	0,059	AGG	0,059	AGG	0,949	AGG	0,949
AGT	0,011	AGT	1,000	AGT	0,000	AGT	0,004	AGT	0,997	AGT	0,297	AGT	0,016	AGT	0,021	AGT	0,021	AGT	0,021	AGT	0,113	AGT	0,113
ATA	1,000	ATA	1,000	ATA	1,000	ATA	1,000	ATA	0,263	ATA	0,016	ATA	0,002	ATA	0,000	ATA	0,000	ATA	0,000	ATA	0,000	ATA	0,000
ATC	1,000	ATC	0,877	ATC	1,000	ATC	1,000	ATC	0,055	ATC	0,002	ATC	0,000	ATC	0,000	ATC	0,000	ATC	0,000	ATC	0,000	ATC	0,000
ATG	1,000	ATG	0,856	ATG	1,000	ATG	1,000	ATG	0,051	ATG	0,002	ATG	0,000	ATG	0,000	ATG	0,000	ATG	0,000	ATG	0,000	ATG	0,000
ATT	0,001	ATT	0,030	ATT	0,008	ATT	0,013	ATT	0,001	ATT	0,000	ATT	0,000	ATT	0,000	ATT	0,000	ATT	0,000	ATT	0,000	ATT	0,000
GAC	0,802	GAA	0,996	GAA	0,942	GAA	1,000	GAA	0,140	GAA	0,006	GAA	0,000	GAA	0,001	GAA	0,000	GAA	0,000	GAA	0,001	GAA	0,001
GAG	0,996	GAC	0,662	GAC	1,000	GAC	1,000	GAC	0,039	GAC	0,003	GAC	0,000	GAC	0,000	GAC	0,000	GAC	0,000	GAC	0,000	GAC	0,000
GAT	1,000	GAT	0,974	GAG	0,688	GAG	0,900	GAG	0,958	GAG	0,283	GAG	0,411	GAG	0,914	GAG	0,001	GAT	0,001	GAT	0,009	GAT	0,009
GCA	0,942	GCA	0,688	GAT	1,000	GAT	1,000	GAT	0,098	GAT	0,004	GAT	0,001	GAT	0,000	GAT	0,000	GAT	0,000	GAT	0,009	GAT	0,009
GCC	0,700	GCC	0,400	GCC	0,995	GCA	1,000	GCA	0,033	GCA	0,001	GCA	0,000	GCA	0,000	GCA	0,000	GCA	0,000	GCA	0,000	GCA	0,000
GCG	1,000	GCG	0,900	GCG	1,000	GCC	0,973	GCC	0,015	GCC	0,001	GCC	0,002	GCC	0,008	GCC	0,008	GCG	0,000	GCG	0,000	GCG	0,000
GCT	0,073	GCT	0,091	GCT	0,476	GCT	0,351	GCT	0,059	GCG	0,002	GCT	0,000	GCT	0,000	GCT	0,000	GCT	0,000	GCT	0,000	GCT	0,000
GGA	0,140	GGA	0,958	GGA	0,033	GGA	0,059	GCT	0,002	GCT	0,000	GCT	0,000	GCT	0,000	GCT	0,000	GCT	0,000	GCT	0,000	GCT	0,000
GGC	0,079	GGC	0,675	GGC	0,025	GGC	0,037	GGC	1,000	GGA	1,000	GGA	1,000	GGA	1,000	GGA	1,000	GGA	1,000	GGA	1,000	GGA	1,000
GGG	0,006	GGG	0,283	GGG	0,001	GGG	0,002	GGG	1,000	GGC	1,000	GGC	1,000	GGC	1,000	GGC	1,000	GGC	1,000	GGC	0,991	GGC	0,991
GGT	0,080	GGT	0,821	GGT	0,021	GGT	0,034	GGT	1,000	GGT	1,000	GGT	1,000	GGG	1,000	GGG	1,000	GGG	1,000	GGG	0,722	GGG	0,722
GTA	0,000	GTA	0,411	GTA	0,000	GTA	0,000	GTA	1,000	GTA	1,000	GTA	1,000	GGT	1,000	GGT	1,000	GGT	1,000	GGT	1,000	GGT	1,000
GTC	0,406	GTC	1,000	GTC	0,005	GTC	0,119	GTC	0,557	GTC	0,046	GTC	0,000	GTC	0,724	GTC	0,000	GTC	0,000	GTC	0,724	GTC	0,724
GTG	0,001	GTG	0,914	GTG	0,000	GTG	0,000	GTG	1,000	GTG	1,000	GTG	0,722	GTG	0,724	GTG	0,724	GTG	0,724	GTG	0,004	GTG	0,004
GTT	1,000	GTT	1,000	GTT	0,834	GTT	0,973	GTT	0,866	GTT	0,170	GTT	0,249	GTT	0,746	GTT	0,746	GTT	0,746	GTT	0,746	GTT	0,746
GAC	AAA	0,013	GAT	AAA	0,042	GCC	AAA	0,995	GCT	AAA	1,000	GGC	AAA	0,001	GGT	AAA	0,001	GTC	AAA	0,000	GTT	AAA	0,030
AAC	0,928	AAC	0,990	AAC	1,000	AAC	0,860	AAC	0,014	AAC	0,011	AAC	0,000	AAC	0,542	AAC	0,542	AAC	0,542	AAC	0,542	AAC	0,542
AAG	1,000	AAG	1,000	AAG	0,952	AAG	0,257	AAG	0,038	AAG	0,034	AAG	0,034	AAG	0,962	AAG	0,962	AAG	0,962	AAG	0,962	AAG	0,962
AAT	0,370	AAT	0,977	AAT	1,000	AAT	0,745	AAT	0,016	AAT	0,013	AAT	0,000	AAT	0,509	AAT	0,509	AAT	0,509	AAT	0,509	AAT	0,509
ACA	0,809	ACA	1,000	ACA	0,868	ACA	0,143	ACA	0,044	ACA	0,040	ACA	0,000	ACA	0,955	ACA	0,955	ACA	0,955	ACA	0,955	ACA	0,955
ACC	0,728	ACC	0,852	ACC	1,000	ACC	1,000	ACC	0,006	ACC	0,004	ACC	0,009	ACC	0,325	ACC	0,325	ACC	0,325	ACC	0,325	ACC	0,325
ACG	1,000	ACG	1,000	ACG	1,000	ACG	0,977	ACG	0,015	ACG	0,013	ACG	0,095	ACG	0,768	ACG	0,768	ACG	0,768	ACG	0,768	ACG	0,768
ACT	0,224	ACT	0,543	ACT	1,000	ACT	1,000	ACT	0,005	ACT	0,003	ACT	0,001	ACT	0,195	ACT	0,195	ACT	0,195	ACT	0,195	ACT	0,195
AGA	0,168	AGA	0,647	AGA	0,122	AGA	0,009	AGA	0,744	AGA	0,882	AGA	1,000	AGA	1,000	AGA	1,000	AGA	1,000	AGA	1,000	AGA	1,000
AGC	0,046	AGC	0,340	AGC	0,059	AGC	0,002	AGC	0,790	AGC	0,917	AGC	0,997	AGC	1,000	AGC	1,000	AGC	1,000	AGC	1,000	AGC	1,000
AGG	0,000	AGG	0,142	AGG	0,046	AGG	0,001	AGG	0,598	AGG	0,731	AGG	0,794	AGG	1,000	AGG	1,000	AGG	1,000	AGG	1,000	AGG	1,000
AGT	0,000	AGT	0,047	AGT	0,023	AGT	0,000	AGT	0,825	AGT	0,937	AGT	0,143	AGT	0,989	AGT	0,989	AGT	0,989	AGT	0,989	AGT	0,989
ATA	1,000	ATA	1,000	ATA	0,956	ATA	0,483	ATA	0,116	ATA	0,146	ATA	0,957	ATA	1,000	ATA	1,000	ATA	1,000	ATA	1,000	ATA	1,000
ATC	1,000	ATC	1,000	ATC	0,966	ATC	0,306	ATC	0,036	ATC	0,033	ATC	0,058	ATC	0,962	ATC	0,962	ATC	0,962	ATC	0,962	ATC	0,962
ATG	1,000	ATG	1,000	ATG	0,975	ATG	0,339	ATG	0,034	ATG	0,030	ATG	0,049	ATG	0,951	ATG	0,951	ATG	0,951	ATG	0,951	ATG	0,951
ATT	0,002	ATT	0,049	ATT	0,999	ATT	1,000	ATT	0,002	ATT	0,001	ATT	0,000	ATT	0,041	ATT	0,041	ATT	0,041	ATT	0,041	ATT	0,041
GAA	0,802	GAA	1,000	GAA	0,700	GAA	0,073	GAA	0,079	GAA	0,080	GAA	0,406	GAA	1,000	GAA	1,000	GAA	1,000	GAA	1,000	GAA	1,000
GAG	0,662	GAC	1,000	GAC	0,975	GAC	0,288	GAC	0,030	GAC	0,026	GAC	0,000	GAC	0,809	GAC	0,809	GAC	0,809	GAC	0,809	GAC	0,809
GAT	1,000	GAG	0,974	GAG	0,400	GAG	0,091	GAG	0,675	GAG	0,821	GAG	1,000	GAG	1,000	GAT	1,000	GAT	1,000	GAT	1,000	GAT	1,000
GCA	1,000	GCA	1,000	GAT	0,970	GAT	0,414	GAT	0,052	GAT	0,053	GAT	0,484	GAT	0,997	GAT	0,997	GAT	0,997	GAT	0,997	GAT	0,997
GCC	0,975	GCC	0,970	GCA	0,995	GCA	0,476	GCA	0,025	GCA	0,021	GCA	0,005	GCA	0,834	GCA	0,834	GCC	0,005	GCC	0,834	GCC	0,834
GCG	1,000	GCG	1,000	GCG	0,973	GCC	1,000	GCC	0,008	GCC	0,008	GCC	0,125	GCC	0,530	GCC	0,530	GCG	0,008	GCG	0,530	GCG	0,530
GCT	0,288	GCT	0,414	GCT	1,000	GCT	0,351	GCG	0,037	GCG	0,034	GCG	0,119	GCG	0,973	GCG	0,973	GCT	0,034	GCT	0,973	GCT	0,973
GGA	0,039	GGA	0,098	GGA	0,015	GGA	0,002	GCT	0,002	GCT	0,001	GCT	0,005	GCT	0,132	GCT	0,132	GGA	0,005	GGA	0,132	GGA	0,132
GGC	0,030	GGC	0,052	GGC	0,008	GGC	0,002	GGA	1,000	GGA	1,000	GGA	0,557	GGA	0,866	GGA	0,866	GGC	0,557	GGC	0,866	GGC	0,866
GGG	0,003	GGG	0,004	GGG	0,001	GGG	0,000	GGG	1,000	GGC	1,000	GGC	0,296	GGC	0,515	GGC	0,515	GGG	0,296	GGG	0,515	GGG	0,515
GGT	0,026	GGT	0,053	GGT	0,008	GGT	0,001	GGT	1,000	GGG	1,000	GGG	0,046	GGG	0,170	GGG	0,170	GGT	0,046	GGT	0,170	GGT	0,170
GTA	0,000	GTA	0,001	GTA	0,002	GTA	0,000	GTA	1,000	GTA	1,000	GTA	0,350	GTA	0,660	GTA	0,660	GTA	0,350	GTA	0,660	GTA	0,660
GTC	0,000	GTC	0,484	GTC	0,125	GTC	0,005	GTC	0,296	GTC	0,350	GTA	0,000	GTA	0,249	GTA	0,249	GTC	0,350	GTC	0,249	GTC	0,249
GTG	0,000	GTG	0,009	GTG	0,008	GTG	0,000	GTG	0,991	GTG	1,000	GTG	0,004	GTG	1,000	GTG	1,000	GTG	0,004	GTG	1,000	GTG	1,000
GTT	0,809	GTT	0,997	GTT	0,530	GTT	0,132	GTT	0,515	GTT													

# T7 promóter eredményeihez kapcsolódó p értékek

(0,000 jelentése <0,001)

41. ábra

A	C	0,000
G		0,000
T		0,000
C	A	0,000
G		0,000
T		0,000
G	A	0,000
C		0,000
T		0,000
T	A	0,000
C		0,000
G		0,000

42. ábra

AA	AC	0,000	CA	AA	0,000	GA	AA	0,000	TA	AA	0,026
AG	0,000		AC	0,000		AC	0,001		AC	0,000	
AT	0,997		AG	0,000		AG	0,441		AG	0,000	
CA	0,000		AT	0,000		AT	0,000		AT	0,351	
CC	0,000		CC	0,601		CA	0,000		CA	0,000	
CG	0,000		CG	0,005		CC	0,000		CC	0,000	
CT	0,172		CT	0,092		CG	0,001		CG	0,000	
GA	0,000		GA	0,000		CT	0,000		CT	0,005	
GC	0,000		GC	0,000		GC	0,002		GA	0,000	
GG	0,000		GG	0,000		GG	0,003		GC	0,000	
GT	0,000		GT	0,000		GT	0,034		GG	0,000	
TA	0,026		TA	0,000		TA	0,000		GT	0,000	
TC	1,000		TC	0,000		TC	0,000		TC	0,112	
TG	0,000		TG	0,000		TG	0,001		TG	0,000	
TT	0,000		TT	0,000		TT	0,000		TT	0,001	
AC	AA	0,000	CC	AA	0,000	GC	AA	0,000	TC	AA	1,000
AG	0,000		AC	0,000		AC	1,000		AC	0,000	
AT	0,000		AG	0,000		AG	0,000		AG	0,000	
CA	0,000		AT	0,000		AT	0,000		AT	1,000	
CC	0,000		CA	0,601		CA	0,000		CA	0,000	
CG	0,110		CG	0,000		CC	0,000		CC	0,000	
CT	0,000		CT	0,443		CG	0,003		CG	0,000	
GA	0,001		GA	0,000		CT	0,000		CT	0,191	
GC	1,000		GC	0,000		GA	0,002		GA	0,000	
GG	0,986		GG	0,000		GG	0,929		GC	0,000	
GT	0,000		GT	0,000		GT	0,000		GG	0,000	
TA	0,000		TA	0,000		TA	0,000		GT	0,000	
TC	0,000		TC	0,000		TC	0,000		TA	0,112	
TG	1,000		TG	0,000		TG	0,999		TG	0,000	
TT	0,000		TT	0,000		TT	0,000		TT	0,001	
AG	AA	0,000	CG	AA	0,000	GG	AA	0,000	TG	AA	0,000
AC	0,000		AC	0,110		AC	0,986		AC	1,000	
AT	0,000		AG	0,000		AG	0,000		AG	0,000	
CA	0,000		AT	0,000		AT	0,000		AT	0,000	
CC	0,000		CA	0,005		CA	0,000		CA	0,000	
CG	0,000		CC	0,000		CC	0,000		CC	0,000	
CT	0,000		CT	0,000		CG	0,020		CG	0,081	
GA	0,441		GA	0,001		CT	0,000		CT	0,000	
GC	0,000		GC	0,003		GA	0,003		GA	0,001	
GG	0,000		GG	0,020		GC	0,929		GC	0,999	
GT	0,014		GT	0,000		GT	0,013		GG	0,846	
TA	0,000		TA	0,000		TA	0,000		GT	0,000	
TC	0,000		TC	0,000		TC	0,000		TA	0,000	
TG	0,000		TG	0,081		TG	0,846		TC	0,000	
TT	0,000		TT	0,000		TT	0,000		TT	0,000	
AT	AA	0,997	CT	AA	0,172	GT	AA	0,000	TT	AA	0,000
AC	0,000		AC	0,000		AC	0,000		AC	0,000	
AG	0,000		AG	0,000		AG	0,014		AG	0,000	
CA	0,000		AT	0,059		AT	0,000		AT	0,001	
CC	0,000		CA	0,092		CA	0,000		CA	0,000	
CG	0,000		CC	0,443		CC	0,000		CC	0,000	
CT	0,059		CG	0,000		CG	0,000		CG	0,000	
GA	0,000		GA	0,000		CT	0,000		CT	0,000	
C	0,000		GC	0,000		GA	0,034		GA	0,000	
GG	0,000		GG	0,000		GC	0,000		GC	0,000	
GT	0,000		GT	0,000		GG	0,013		GG	0,000	
TA	0,351		TA	0,005		TA	0,000		GT	0,000	
TC	1,000		TC	0,191		TC	0,000		TA	0,001	
TG	0,000		TG	0,000		TG	0,000		TC	0,001	
TT	0,001		TT	0,000		TT	0,000		TG	0,000	

43. ábra

GAA	GAC	0,000	GCA	GAA	0,000	GGA	GAA	0,000	GTA	GAA	0,796
GAG	0,002		GAC	0,998		GAC	1,000		GAC	0,007	
GAT	0,012		GAG	0,279		GAG	1,000		GAG	0,510	
GCA	0,000		GAT	0,000		GAT	0,000		GAT	0,000	
GCC	0,000		GCC	0,000		GCA	0,916		GCA	0,000	
GCG	0,000		GCG	0,065		GCC	0,000		GCC	0,000	
GCT	0,000		GCT	0,997		GCG	0,000		GCG	0,000	
GGA	0,000		GGA	0,916		GCT	1,000		GCT	0,010	
GGC	0,000		GGC	0,237		GGC	0,001		GGA	0,053	
GGG	0,000		GGG	0,000		GGG	0,000		GGC	0,000	
GGT	0,000		GGT	1,000		GGT	0,999		GGG	0,000	
GTA	0,796		GTA	0,000		GTA	0,053		GGT	0,001	
GTC	0,000		GTC	0,116		GTC	0,000		GTC	0,000	
GTG	0,000		GTG	1,000		GTG	0,998		GTG	0,001	
GTT	0,000		GTT	0,998		GTT	1,000		GTT	0,008	
GAC	GAA	0,000	GCC	GAA	0,000	GGC	GAA	0,000	GTC	GAA	0,000
GAG	0,959		GAC	0,000		GAC	0,008		GAC	0,003	
GAT	0,000		GAG	0,000		GAG	0,000		GAG	0,000	
GCA	0,998		GAT	0,000		GAT	0,000		GAT	0,000	
GCC	0,000		GCA	0,000		GCA	0,237		GCA	0,116	
GCG	0,001		GCG	0,428		GCC	0,148		GCC	0,289	
GCT	1,000		GCT	0,000		GCG	1,000		GCG	1,000	
GGA	1,000		GGA	0,000		GCT	0,006		GCT	0,002	
GGC	0,008		GGC	0,148		GGA	0,001		GGA	0,000	
GGG	0,000		GGG	1,000		GGG	0,073		GGC	1,000	
GGT	1,000		GGT	0,000		GGT	0,036		GGG	0,160	
GTA	0,007		GTA	0,000		GTA	0,000		GGT	0,014	
GTC	0,003		GTC	0,289		GTC	1,000		GTA	0,000	
GTG	1,000		GTG	0,000		GTG	0,054		GTG	0,021	
GTT	1,000		GTT	0,000		GTT	0,007		GTT	0,002	
GAG	GAA	0,002	GCG	GAA	0,000	GGG	GAA	0,000	GTG	GAA	0,000
GAC	0,959		GAC	0,001		GAC	0,000		GAC	1,000	
GAT	0,000		GAG	0,000		GAG	0,000		GAG	0,677	
GCA	0,279		GAT	0,000		GAT	0,000		GAT	0,000	
GCC	0,000		GCA	0,065		GCA	0,000		GCA	1,000	
GCG	0,000		GCC	0,428		GCC	1,000		GCC	0,000	
GCT	0,975		GCT	0,001		GCG	0,259		GCG	0,010	
GGA	1,000		GGA	0,000		GCT	0,000		GCT	1,000	
GGC	0,000		GGC	1,000		GGA	0,000		GGA	0,998	
GGG	0,000		GGG	0,259		GGC	0,073		GGC	0,054	
GGT	0,769		GGT	0,006		GGT	0,000		GGG	0,000	
GTA	0,510		GTA	0,000		GTA	0,000		GGT	1,000	
GTC	0,000		GTC	1,000		GTC	0,160		GTA	0,001	
GTG	0,677		GTG	0,010		GTG	0,000		GTC	0,021	
GTT	0,965		GTT	0,001		GTT	0,000		GTT	1,000	
GAT	GAA	0,012	GCT	GAA	0,000	GGT	GAA	0,000	GTT	GAA	0,000
GAC	0,000		GAC	1,000		GAC	1,000		GAC	1,000	
GAG	0,000		GAG	0,975		GAG	0,769		GAG	0,965	
GCA	0,000		GAT	0,000		GAT	0,000		GAT	0,000	
GCC	0,000		GCA	0,997		GCA	1,000		GCA	0,998	
GCG	0,000		GCC	0,000		GCC	0,000		GCC	0,000	
GCT	0,000		GCG	0,001		GCG	0,006		GCG	0,001	
GGA	0,000		GGA	1,000		GCT	1,000		GCT	1,000	
GGC	0,000		GGC	0,006		GGA	0,999		GGA	1,000	
GGG	0,000		GGG	0,000		GGC	0,036		GGC	0,007	
GGT	0,000		GGT	1,000		GGG	0,000		GGG	0,000	
GTA	0,000		GTA	0,010		GTA	0,001		GGT	1,000	
GTC	0,000		GTC	0,002		GTC	0,014		GTA	0,008	
GTG	0,000		GTG	1,000		GTG	1,000		GTC	0,002	
GTT	0,000		GTT	1,000		GTT	1,000		GTG	1,000	

44. ábra

GGAA	GGAC	0,023	GGCA	GGAA	0,007	GGGA	GGAA	0,000	GGTA	GGAA	0,164
GGAG	0,047		GGAC	0,968		GGAC	0,005		GGAC	0,093	
GGAT	0,069		GGAG	0,999		GGAG	0,178		GGAG	0,787	
GGCA	0,007		GGAT	0,539		GGAT	0,000		GGAT	0,678	
GGCC	0,000		GGCC	0,109		GGCA	0,729		GGCA	0,168	
GGCG	0,001		GGCG	0,998		GGCC	0,674		GGCC	0,000	
GGCT	0,121		GGCT	0,806		GGCG	0,996		GGCG	0,006	
GGGA	0,000		GGGA	0,729		GGCT	0,017		GGCT	0,998	
GGGC	0,000		GGGC	0,040		GGGC	0,134		GGGA	0,000	
GGGG	0,000		GGGG	0,009		GGGG	0,012		GGGC	0,000	
GGGT	0,001		GGGT	0,283		GGGT	0,997		GGGG	0,000	
GGTA	0,164		GGTA	0,168		GGTA	0,000		GGGT	0,000	
GGTC	0,003		GGTC	1,000		GGTC	0,508		GGTC	0,000	
GGTG	0,004		GGTG	1,000		GGTG	0,396		GGTG	0,008	
GGTT	1,000		GGTT	0,476		GGTT	0,112		GGTT	0,991	
GGAC	GGAA	0,023	GGCC	GGAA	0,000	GGGC	GGAA	0,000	GGTC	GGAA	0,003
GGAG	1,000		GGAC	0,000		GGAC	0,000		GGAC	0,029	
GGAT	0,928		GGAG	0,021		GGAG	0,010		GGAG	0,673	
GGCA	0,968		GGAT	0,000		GGAT	0,000		GGAT	0,000	
GGCC	0,000		GGCA	0,109		GGCA	0,040		GGCA	1,000	
GGCG	0,182		GGCG	0,226		GGCC	0,638		GGCC	0,000	
GGCT	0,999		GGCT	0,002		GGCG	0,056		GGCG	1,000	
GGGA	0,005		GGGA	0,674		GGCT	0,001		GGCT	0,127	
GGGC	0,000		GGGC	0,638		GGGA	0,134		GGGA	0,508	
GGGG	0,000		GGGG	0,020		GGGG	0,127		GGGC	0,000	
GGGT	0,000		GGGT	0,644		GGGT	0,000		GGGG	0,000	
GGTA	0,093		GGTA	0,000		GGTA	0,000		GGGT	0,000	
GGTC	0,029		GGTC	0,000		GGTC	0,000		GGTA	0,000	
GGTG	0,621		GGTG	0,005		GGTG	0,002		GGTG	0,998	
GGTT	0,753		GGTT	0,047		GGTT	0,031		GGTT	0,258	
GGAG	GGAA	0,047	GGCG	GGAA	0,001	GGGG	GGAA	0,000	GGTG	GGAA	0,004
GGAC	1,000		GGAC	0,182		GGAC	0,000		GGAC	0,621	
GGAT	0,998		GGAG	0,665		GGAG	0,002		GGAG	0,977	
GGCA	0,999		GGAT	0,024		GGAT	0,000		GGAT	0,068	
GGCC	0,021		GGCA	0,998		GGCA	0,009		GGCA	1,000	
GGCG	0,665		GGCC	0,226		GGCC	0,020		GGCC	0,005	
GGCT	1,000		GGCT	0,144		GGCG	0,008		GGCG	0,993	
GGGA	0,178		GGGA	0,996		GGCT	0,000		GGCT	0,459	
GGGC	0,010		GGGC	0,056		GGGA	0,012		GGGA	0,396	
GGGG	0,002		GGGG	0,008		GGGC	0,127		GGGC	0,002	
GGGT	0,057		GGGT	0,629		GGGT	0,000		GGGG	0,000	
GGTA	0,787		GGTA	0,006		GGTA	0,000		GGGT	0,031	
GGTC	0,673		GGTC	1,000		GGTC	0,000		GGTA	0,008	
GGTG	0,977		GGTG	0,993		GGTG	0,000		GGTC	0,998	
GGTT	0,814		GGTT	0,219		GGTT	0,017		GGTT	0,384	
GGAT	GGAA	0,069	GGCT	GGAA	0,121	GGGT	GGAA	0,001	GGTT	GGAA	1,000
GGAC	0,928		GGAC	0,999		GGAC	0,000		GGAC	0,753	
GGAG	0,998		GGAG	1,000		GGAG	0,057		GGAG	0,814	
GGCA	0,539		GGAT	1,000		GGAT	0,000		GGAT	0,929	
GGCC	0,000		GGCA	0,806		GGCA	0,283		GGCA	0,476	
GGCG	0,024		GGCC	0,002		GGCC	0,644		GGCC	0,047	
GGCT	1,000		GGCG	0,144		GGCG	0,629		GGCG	0,219	
GGGA	0,000		GGGA	0,017		GGCT	0,006		GGCT	0,957	
GGGC	0,000		GGGC	0,001		GGGA	0,997		GGGA	0,112	
GGGG	0,000		GGGG	0,000		GGGC	0,000		GGGC	0,031	
GGGT	0,000		GGGT	0,006		GGGG	0,000		GGGG	0,017	
GGTA	0,678		GGTA	0,998		GGTA	0,000		GGGT	0,075	
GGTC	0,000		GGTC	0,127		GGTC	0,000		GGTA	0,991	
GGTG	0,068		GGTG	0,459		GGTG	0,031		GGTC	0,258	
GGTT	0,929		GGTT	0,957		GGTT	0,075		GGTG	0,384	

# U6T7 promóter eredményeihez kapcsolódó p értékek

(0,000 jelentése <0,001)

48. ábra

A	C	0,000
G		0,001
T		0,999
C	A	0,000
G		0,000
T		0,000
G	A	0,001
C		0,000
T		0,000
T	A	0,999
C		0,000
G		0,000

49. ábra

AA	AC	0,998	CA	AA	1,000	GA	AA	0,011	TA	AA	0,000
AG		0,000	AC		1,000	AC		0,016	AC		0,000
AT		0,246	AG		0,000	AG		0,629	AG		0,988
CA		1,000	AT		0,301	AT		0,164	AT		0,000
CC		0,220	CC		0,196	CA		0,012	CA		0,000
CG		0,643	CG		0,541	CC		0,000	CC		0,000
CT		0,014	CT		0,011	CG		0,003	CG		0,000
GA		0,011	GA		0,012	CT		0,000	CT		0,000
GC		0,552	GC		0,801	GC		0,022	GA		0,962
GG		0,000	GG		0,000	GG		0,054	GC		0,000
GT		0,000	GT		0,000	GT		0,974	GG		0,094
TA		0,000	TA		0,000	TA		0,962	GT		1,000
TC		0,172	TC		0,234	TC		0,082	TC		0,000
TG		0,000	TG		0,000	TG		0,141	TG		0,000
TT		0,996	TT		0,989	TT		0,004	TT		0,000
AC	AA	0,998	CC	AA	0,220	GC	AA	0,552	TC	AA	0,172
AG		0,000	AC		0,141	AC		1,000	AC		0,562
AT		0,543	AG		0,000	AG		0,000	AG		0,000
CA		1,000	AT		0,020	AT		0,625	AT		1,000
CC		0,141	CA		0,196	CA		0,801	CA		0,234
CG		0,331	CG		0,746	CC		0,102	CC		0,028
CT		0,006	CT		1,000	CG		0,112	CG		0,020
GA		0,016	GA		0,000	CT		0,004	CT		0,000
GC		1,000	GC		0,102	GA		0,022	GA		0,082
GG		0,000	GG		0,000	GG		0,000	GC		0,631
GT		0,000	GT		0,000	GT		0,000	GG		0,000
TA		0,000	TA		0,000	TA		0,000	GT		0,000
TC		0,562	TC		0,028	TC		0,631	TA		0,000
TG		0,006	TG		0,015	TG		0,000	TG		0,994
TT		0,933	TT		0,701	TT		0,746	TT		0,240
AG	AA	0,000	CG	AA	0,643	GG	AA	0,000	TG	AA	0,000
AC		0,000	AC		0,331	AC		0,000	AC		0,006
AT		0,000	AG		0,000	AG		0,314	AG		0,000
CA		0,000	AT		0,038	AT		0,000	AT		1,000
CC		0,000	CA		0,541	CA		0,000	CA		0,000
CG		0,000	CC		0,746	CC		0,000	CC		0,015
CT		0,000	CT		0,320	CG		0,000	CG		0,001
GA		0,629	GA		0,003	CT		0,000	CT		0,000
GC		0,000	GC		0,112	GA		0,054	GA		0,141
GG		0,314	GG		0,000	GC		0,000	GC		0,000
GT		0,999	GT		0,000	GT		0,180	GG		0,000
TA		0,988	TA		0,000	TA		0,094	GT		0,001
TC		0,000	TC		0,020	TC		0,000	TA		0,000
TG		0,000	TG		0,001	TG		0,000	TC		0,994
TT		0,000	TT		1,000	TT		0,000	TT		0,052
AT	AA	0,246	CT	AA	0,014	GT	AA	0,000	TT	AA	0,996
AC		0,543	AC		0,006	AC		0,000	AC		0,933
AG		0,000	AG		0,000	AG		0,999	AG		0,000
CA		0,301	AT		0,001	AT		0,001	AT		0,209
CC		0,020	CA		0,011	CA		0,000	CA		0,989
CG		0,038	CC		1,000	CC		0,000	CC		0,701
CT		0,001	CG		0,320	CG		0,000	CG		1,000
GA		0,164	GA		0,000	CT		0,000	CT		0,382
C		0,625	GC		0,004	GA		0,974	GA		0,004
GG		0,000	GG		0,000	GC		0,000	GC		0,746
GT		0,001	GT		0,000	GG		0,180	GG		0,000
TA		0,000	TA		0,000	TA		1,000	GT		0,000
TC		1,000	TC		0,000	TC		0,000	TA		0,000
TG		1,000	TG		0,000	TG		0,001	TC		0,240
TT		0,209	TT		0,382	TT		0,000	TG		0,052



# J23119 promóter eredményeihez kapcsolódó p értékek

(0,000 jelentése <0,001)

53. ábra

A	C	0,000
G		0,612
T		0,000
C	A	0,000
G		0,000
T		0,000
G	A	0,612
C		0,000
T		0,000
T	A	0,000
C		0,000
G		0,000

54. ábra

AA	AC	0,000	CA	AA	0,000	GA	AA	1,000	TA	AA	1,000
AG		0,000	AC		0,330	AC		0,000	AC		0,000
AT		0,000	AG		0,000	AG		0,000	AG		0,000
CA		0,000	AT		0,000	AT		0,000	AT		0,000
CC		0,000	CC		0,000	CA		0,000	CA		0,000
CG		0,000	CG		0,983	CC		0,000	CC		0,000
CT		0,000	CT		0,003	CG		0,000	CG		0,000
GA		1,000	GA		0,000	CT		0,000	CT		0,000
GC		0,000	GC		0,033	GC		0,000	GA		0,081
GG		0,396	GG		0,000	GG		0,081	GC		0,000
GT		0,000	GT		0,001	GT		0,000	GG		0,010
TA		1,000	TA		0,000	TA		0,081	GT		0,000
TC		0,000	TC		0,467	TC		0,000	TC		0,000
TG		0,963	TG		0,000	TG		0,432	TG		0,905
TT		0,000	TT		0,482	TT		0,000	TT		0,000
AC	AA	0,000	CC	AA	0,000	GC	AA	0,000	TC	AA	0,000
AG		0,000	AC		0,000	AC		0,738	AC		0,001
AT		0,000	AG		0,000	AG		0,000	AG		0,000
CA		0,330	AT		0,000	AT		0,000	AT		0,000
CC		0,000	CA		0,000	CA		0,033	CA		0,467
CG		1,000	CG		0,000	CC		0,000	CC		0,000
CT		0,001	CT		1,000	CG		0,599	CG		0,185
GA		0,000	GA		0,000	CT		0,000	CT		0,010
GC		0,738	GC		0,000	GA		0,000	GA		0,000
GG		0,000	GG		0,000	GG		0,000	GC		0,001
GT		0,218	GT		0,000	GT		1,000	GG		0,000
TA		0,000	TA		0,000	TA		0,000	GT		0,000
TC		0,001	TC		0,000	TC		0,001	TA		0,000
TG		0,000	TG		0,000	TG		0,000	TG		0,000
TT		0,996	TT		0,000	TT		1,000	TT		0,095
AG	AA	0,000	CG	AA	0,000	GG	AA	0,396	TG	AA	0,963
AC		0,000	AC		1,000	AC		0,000	AC		0,000
AT		0,000	AG		0,000	AG		0,000	AG		0,000
CA		0,000	AT		0,000	AT		0,000	AT		0,000
CC		0,000	CA		0,983	CA		0,000	CA		0,000
CG		0,000	CC		0,000	CC		0,000	CC		0,000
CT		0,000	CT		0,001	CG		0,000	CG		0,000
GA		0,000	GA		0,000	CT		0,000	CT		0,000
GC		0,000	GC		0,599	GA		0,081	GA		0,432
GG		0,000	GG		0,000	GC		0,000	GC		0,000
GT		0,000	GT		0,299	GT		0,000	GG		0,023
TA		0,000	TA		0,000	TA		0,010	GT		0,000
TC		0,000	TC		0,185	TC		0,000	TA		0,905
TG		0,000	TG		0,000	TG		0,023	TC		0,000
TT		0,000	TT		0,966	TT		0,000	TT		0,000
AT	AA	0,000	CT	AA	0,000	GT	AA	0,000	TT	AA	0,000
AC		0,000	AC		0,001	AC		0,218	AC		0,996
AG		0,000	AG		0,000	AG		0,000	AG		0,000
CA		0,000	AT		0,003	AT		0,000	AT		0,000
CC		0,000	CA		0,003	CA		0,001	CA		0,482
CG		0,000	CC		1,000	CC		0,000	CC		0,000
CT		0,003	CG		0,001	CG		0,299	CG		0,966
GA		0,000	GA		0,000	CT		0,000	CT		0,000
C		0,000	GC		0,000	GA		0,000	GA		0,000
GG		0,000	GG		0,000	GC		1,000	GC		1,000
GT		0,000	GT		0,000	GG		0,000	GG		0,000
TA		0,000	TA		0,000	TA		0,000	GT		1,000
TC		0,000	TC		0,010	TC		0,000	TA		0,000
TG		0,000	TG		0,000	TG		0,000	TC		0,095
TT		0,000	TT		0,000	TT		1,000	TG		0,000